An Approach for Obtaining and Extracting Occupancy Patterns in Buildings Based on Mobile Positioning Data

Xuyuan Kang¹, Da Yan¹, Hongsan Sun¹, Yuan Jin¹, Peng Xu²
¹Building Energy Research Centre, Tsinghua University, Beijing, China
²Tongji University, Shanghai, China

Abstract
Occupancy data is an essential input parameter for building energy simulation and has a huge impact on the accuracy of building energy models. Current approaches of obtaining occupancy data usually involve sophisticated occupancy detection devices and algorithms, or time-consuming manual counting work. However, with the advancement of information technology and network service, the amount of data has accumulated sharply. The growing popularity of mobile devices has made it possible to acquire location data within a specific region based on active positioning requests from end-users, offering a new approach to obtain and extract occupancy profiles for different buildings and areas. This research uses clustering methods to extract typical daily and weekly occupancy profiles based on big data of location. Based on clustering analysis, several descriptive indices are introduced to establish occupancy models for different types of buildings. Moreover, comparisons of occupancy profiles between location data and energy codes are conducted to demonstrate the significance of location-data-based occupancy profiles.

Introduction
Heating and cooling load of the building is determined by multiple factors including the insulation of building envelope, the optical properties of the windows, the infiltration of the outside air, the lighting and appliances in buildings and occupant behaviour in buildings. Among all these factors, occupant behaviour contributes most, as it not only performs as an internal heat gain of occupants, but also dominates the active control of appliances and equipment like windows, fans and air-conditioners. Thus, it is very essential to acquire building occupancy data for the design and simulation of building energy systems.

Currently there are a few approaches to acquire such kind of occupancy data. The first approach is based on experience or energy codes. The experience usually comes from the modeler’s common sense or recognition of certain types of buildings, while the occupancy data in energy codes usually orient from large scale investigation conducted years ago for a certain general type of building. For example, ASHRAE Standard 90.1 specified several recommended occupancy entries for certain types of common buildings. The occupancy data from these sources are frequently used in the design or modelling process for its accessibility and convenience, but usually lacks specificity and accuracy.

Another approach for acquiring occupancy data is sensor-based method. PIR sensor is one of the commonly installed sensors in buildings. The PIR sensor counts people entering or leaving the building from the exits and sums up as the total number of occupants in buildings. The PIR sensor is cheap and easy to install, but the accuracy is the main concern, as the repeat record or the missing record will cause significant accumulative error to the total number. Many researches focus on the correcting algorithms to compensate for this flaw, but are still not effective enough.

The third approach is manual counting. This method is being used to acquire occupancy data for energy models of train and airport terminals. This method consumes huge manpower and is very inefficient.

The current approaches of acquiring occupancy data are traditional and may not cater for the need of modelling and design process of the building energy system. With the development of computer technology, Internet connection and data science, massive data along with the data analysis is available under many scenarios, and may benefit the collection and analysis of occupancy data in buildings.

There are researchers that took advantage of this occupancy mega data to investigate the daily occupant behaviour in specific buildings. Gu et al. (2018) uses clustering methods to extract typical daily occupant behaviour patterns in public buildings. Zou et al. (2018) analysed WIFI-based data for occupant behaviour prediction and lighting control. These researches provide insights for mega data analysis in building occupant behaviour research.

This paper utilizes the data from active positioning requests of social network software from Tencent Cooperation on their mobile devices, which is one of the largest and most popular IT company in China. The server collects active positioning data from the users who require positioning service when using Tencent apps within a specific region. The positioning data represents an active positioning service when using Tencent apps within a specific region. The positioning data contains information of occupancy patterns within a specific region for a certain period of time.

In this research, clustering methods are used to extract occupancy patterns from mobile positioning data, and
gives descriptive parameters to represent the features of each pattern. The novelty of this research is listed below:

- Using descriptive parameters to represent different occupancy patterns.
- Illustrating typical occupancy portraits with clustering methods.
- Focusing on both daily occupancy profiles and weekly occupancy profiles.
- Comparing differences on both occupancy patterns and the results of energy performance simulation.

The organization of this paper is clarified as follows. The second part explains the methodology of this research. The third part demonstrates the result of typical occupancy patterns in different types of buildings. The fourth part discusses the difference between positioning-data-based occupancy patterns and code-based occupancy patterns, as well as its influence on building energy modelling. The final part concludes the limitations and future perspectives.

**Methodology**

The aim of this research is to extract both daily and weekly patterns of occupancy in buildings. The methods for both ends are generally identical, with only minor differences. The overall procedure follows 4 steps:

**Step 1: Normalization**

The original positioning data is formatted as total number of occupants in a specific region. The first step is to normalize the raw data to get the occupancy value. Max normalization is used in this step. For daily profiles, the max normalization scales the data to the range of [0,1] by daily maximum value, and can be expressed as follows:

\[
a_i = \frac{x_i}{\max_{i=1,24} \{x_i\}} \quad (1)
\]

In which \( \max_{i=1,24} \{x_i\} \) represents the maximum value within the 24 hours of a day.

While for weekly profiles, the max normalization scales the data to the range of [0,1] by weekly maximum value, and can be expressed as follows:

\[
a_i = \frac{x_i}{\max_{i=1,168} \{x_i\}} \quad (2)
\]

In which \( \max_{i=1,168} \{x_i\} \) represents the maximum value within the 168 hours of a week.

**Step 2: Clustering**

Taking normalized data as input, the clustering method is conducted to extract typical occupancy patterns. K-means method is one of the common clustering methods. This method requires the number of clusters as an input. This input is selected based on Calinski-Harabaz (C-H) Principle. C-H index is defined as the ratio of within-cluster dispersion and the between-cluster dispersion. It can be denoted as:

\[
S(k) = \frac{\text{tr}(B_k) + (m-k)}{\text{tr}(W_k) + (k-1)} \quad (3)
\]

Where \( m \) is the number of samples, \( k \) is the number of clusters, \( B_k \) refers to the between-cluster dispersion matrix, \( W_k \) refers to the within-cluster dispersion matrix, \( tr \) refers to the trace of the matrix. \( S(k) \) is the Calinski-Harabaz index.

It can be seen through the definition of C-H index that greater the C-H value, better the clustering effect. As a result, the number of clusters is that with the greatest C-H value.

After clustering, the temporal distribution of each cluster is analysed. For daily profile, each pattern represents weekdays, weekends or holidays in a week. While for weekly profile, each pattern represents different months in a year. The temporal distribution analysis can help better identify the different occupancy features of different time periods.

**Step 3: Parametric Description**

A set of parameters are introduced to describe the different features of each patterns. Different parameters are introduced for daily and weekly profiles. For daily profiles, the following 3 parameters are introduced:

- Peak period: refers to the period with occupancy greater than 0.7.
- Valley period: refers to the period with occupancy lower than 0.3.
- Peak to valley ratio: refers to the ratio of the maximum occupancy to the minimum occupancy in a day.

The peak and valley period are two essential parameters for the control of building energy systems. The peak to valley ratio, however, is crucial for the selection and matching of the capacity of building energy systems. It is to be noted that 0.7/0.3 is not fixed and may vary by different system configurations.

For weekly profiles, the following 2 parameters are introduced:

- Peak ratio: refers to the ratio of maximum daily peak value to minimum daily peak value.
- Daily total occupant ratio: refers to the ratio of maximum daily total number of occupants to minimum daily total number of occupants.

The two parameters of weekly profiles are both essential indicators for the control of building energy systems within a scale of week.

Based on the method mentioned above, the framework of this research is shown below:

![Figure 1: Technical route of occupancy research based on mobile positioning data.](image-url)
Case study
In this research, mobile positioning data from 16 public buildings is used in the analysis of occupancy patterns. These buildings include transport, commercial and hospital, locating in Beijing and Shanghai in China, covering a time span of approximately 2 years from Dec 1st, 2015 to Sept 6th, 2017. The details of the case study buildings are shown below.

Table 1: Details of case study buildings.

<table>
<thead>
<tr>
<th>No.</th>
<th>Building</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Railway Station 1 (Highspeed)</td>
<td>Beijing</td>
</tr>
<tr>
<td>2</td>
<td>Railway Station 2 (Highspeed)</td>
<td>Beijing</td>
</tr>
<tr>
<td>3</td>
<td>Railway Station 3 (Normal Speed)</td>
<td>Beijing</td>
</tr>
<tr>
<td>4</td>
<td>Railway Station 4 (Highspeed)</td>
<td>Shanghai</td>
</tr>
<tr>
<td>5</td>
<td>Airport 1</td>
<td>Beijing</td>
</tr>
<tr>
<td>6</td>
<td>Airport 2</td>
<td>Shanghai</td>
</tr>
<tr>
<td>7</td>
<td>Airport 3</td>
<td>Shanghai</td>
</tr>
<tr>
<td>8</td>
<td>Commercial Complex 1</td>
<td>Beijing</td>
</tr>
<tr>
<td>9</td>
<td>Commercial Complex 2</td>
<td>Beijing</td>
</tr>
<tr>
<td>10</td>
<td>Commercial Complex 3</td>
<td>Beijing</td>
</tr>
<tr>
<td>11</td>
<td>Commercial Complex 4</td>
<td>Shanghai</td>
</tr>
<tr>
<td>12</td>
<td>Commercial Complex 5</td>
<td>Shanghai</td>
</tr>
<tr>
<td>13</td>
<td>Commercial Complex 6</td>
<td>Shanghai</td>
</tr>
<tr>
<td>14</td>
<td>Commercial Complex 7</td>
<td>Shanghai</td>
</tr>
<tr>
<td>15</td>
<td>Commercial Complex 8</td>
<td>Shanghai</td>
</tr>
<tr>
<td>16</td>
<td>Hospital 1</td>
<td>Beijing</td>
</tr>
</tbody>
</table>

Daily occupancy profiles
- Railway Station
Railway Station 3 is taken as the example of Railway Station type. Based on K-means clustering method and Calinski-Harabaz principle, three typical daily occupancy patterns are extracted. Meanwhile, the temporal distribution of each pattern is plotted in a radar graph. The results are shown below:

![Figure 2: Typical daily occupancy patterns and temporal distribution in Railway Station 3.](image)

Generally, for each pattern, there’s a peak from around 11:00 to 24:00, and a valley from around 1:00 to 6:00. There are slight difference of peak and valley period and significant difference of peak valley ratio for each pattern. The details of the descriptive parameters are shown in the table below.

Table 2: Descriptive parameters of typical daily patterns for Railway Station 3.

<table>
<thead>
<tr>
<th>Pattern No.</th>
<th>Peak Period</th>
<th>Valley Period</th>
<th>Peak Valley Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10:00-24:00</td>
<td>1:00-6:00</td>
<td>8.44</td>
</tr>
<tr>
<td>1</td>
<td>11:00-24:00</td>
<td>1:00-6:00</td>
<td>10.82</td>
</tr>
<tr>
<td>2</td>
<td>11:00-24:00</td>
<td>4:00-6:00</td>
<td>3.69</td>
</tr>
</tbody>
</table>

As we can see through the temporal distribution, Pattern 0 represents the occupancy patterns from Monday to Thursday as well as Saturday, Pattern 1 represents the occupancy patterns of Friday and Sunday, showing a greater peak valley ratio, which is the result of short trips on weekends. Pattern 2 represents the occupancy patterns of holidays, which shows a smaller difference of peak and valley period.

- Commercial Building
Commercial complex 1 is chosen as the case of the Commercial Building type. Commercial complex 1 is a large commercial complex in Beijing. Similarly, based on methods mentioned above, two typical daily occupancy patterns as well as temporal distribution are shown below.

![Figure 3: Typical daily occupancy patterns and temporal distribution in Commercial complex 1.](image)

It is to be noted that commercial complexes have opening hours within a day. As a result, it is senseless to use valley period or peak valley ratio as the descriptive parameters. Thus, peak period is the only parameter shown in the graph and enlisted in the table below.

From the temporal distribution, it can be seen clearly that Pattern 0 represents the occupancy of weekends and
holidays, while Pattern 1 represents the occupancy of weekdays.

Table 3: Descriptive parameters of typical daily patterns for Commercial complex 1.

<table>
<thead>
<tr>
<th>Pattern No.</th>
<th>Peak Period</th>
<th>Valley Period</th>
<th>Peak Valley Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12:00-20:00</td>
<td>00:00-6:00</td>
<td>10.2</td>
</tr>
<tr>
<td>1</td>
<td>9:00-16:00</td>
<td>23:00-6:00</td>
<td>15.1</td>
</tr>
</tbody>
</table>

The two typical daily occupancy patterns are of great similarity, no matter from the figure or the descriptive parameters. The only difference between the two patterns is that, Pattern 1, which represents the occupancy of weekdays, exists a “noon peak” and a “evening peak”. This indicates commercial complex’s dining service for people during workdays.

- **Hospital**

Hospital 1 is located in Beijing. From the clustering analysis, two typical daily occupancy patterns are extracted from the positioning data. The results are shown below.

![Figure 4: Typical daily occupancy patterns and temporal distribution in Hospital 1.](https://example.com/figure4.png)

From the temporal distribution, the two patterns can explicitly distinguish from each other by weekdays and weekends. Pattern 0 represents weekends and holidays, while Pattern 1 represents weekdays.

For the occupancy patterns in hospital, the peak period usually distributes from 9:00 to 16:00, while the valley usually distributes from 0:00 to 6:00. The main difference between weekday and weekend patterns is the peak valley ratio, with a higher peak in weekday pattern, suggesting more people coming for outpatient service on weekdays. The detailed descriptive parameters are shown in the table below.

Table 4: Descriptive parameters of typical daily patterns for Hospital 1.

- **Typical Daily Occupancy Pattern Portrait**

In this sub-section, the occupancy patterns of all 16 buildings are portrayed in order to display a general feature of all buildings. In order to portrait this, occupancy data of all buildings are clustered into one single cluster using K-means method, and the result represents the occupancy pattern of specific building. The results are shown in Figure 5.

From the portraits, we can see that the occupancy patterns of highspeed railway stations (Railway Station 1,2,4) agree well with each other, while for normal speed station (Railway Station 3), there are still a significant amount of people in station during night time. Similarly, the airports also show similar occupancy features.

The occupancy patterns of different commercial complexes vary in details. That’s because the occupancy for commercial complexes is largely determined by their opening hours. And the peak usually appears during daytime.

![Figure 5: Typical Daily Occupancy Pattern Portrait of 16 case buildings.](https://example.com/figure5.png)

- **Weekly occupancy profiles**

Following the same procedure of daily occupancy profile analysis, the weekly features of occupancy in 16 case buildings are also analyzed. In this section, the same 3 buildings are chosen to demonstrate the results.

- **Railway Station**

The typical weekly patterns of Railway Station 3 and their temporal distribution are shown below. Two typical patterns are extracted from the positioning data. From the patterns of weekly profiles, it’s easy to infer that number of occupants reaches its peak on Fridays.
From the temporal distribution, Pattern 0 represents the patterns from January to September, while Pattern 1 represents the patterns from October to December. Pattern 1 has a higher Peak Ratio and Daily Total Occupant Ratio, which suggests more people traveling on Fridays.

The result of temporal distribution suggests that Pattern 0 mainly represents the occupancy from September to November, while Pattern 1 mainly represents the occupancy from June to August. There are no significant differences for other months. Although the peak both happens on Fridays, in Pattern 0 there are more occupants on weekends than on weekdays, while in Pattern 1 more occupants show up on weekdays. This might have connections to summer holidays.

The detailed parameters are shown in the table below.

<table>
<thead>
<tr>
<th>Pattern No.</th>
<th>Peak Ratio</th>
<th>Daily Total Occupant Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.23</td>
<td>1.18</td>
</tr>
<tr>
<td>1</td>
<td>1.50</td>
<td>1.34</td>
</tr>
</tbody>
</table>

**Commercial Building**

There are two typical weekly patterns in Commercial Complex 1. The patterns and the temporal distribution are shown below.

- Hospital

In Hospital 1, two typical weekly occupancy patterns are extracted from the positioning data. Below are the results of the patterns and their temporal distribution.

**Table 6: Descriptive parameters of typical weekly patterns for Commercial Complex 1.**

<table>
<thead>
<tr>
<th>Pattern No.</th>
<th>Peak Ratio</th>
<th>Daily Total Occupant Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.21</td>
<td>1.18</td>
</tr>
<tr>
<td>1</td>
<td>1.18</td>
<td>1.23</td>
</tr>
</tbody>
</table>

**Figure 7: Typical weekly occupancy patterns and temporal distribution in Commercial Complex 1.**

As we can see through the result of temporal distribution, Pattern 0 represents the occupancy features from February to September, Pattern 1 represents the occupancy features from November to January, while Pattern 2 only represents two consecutive weeks in October. It is believed that Pattern 2 is an abnormal operating condition, and should not be further considered as the frequent operation schedule.

The result suggests that Pattern 0 has a higher Peak Ratio and Daily Total Occupant Ratio than Pattern 1, which indicates more people on weekdays or fewer people on weekends in Pattern 0.

The details of the parameters are shown below.

**Figure 8: Typical weekly occupancy patterns and temporal distribution in Hospital 1.**
Table 7: Descriptive parameters of typical weekly patterns for Hospital 1.

<table>
<thead>
<tr>
<th>Pattern No.</th>
<th>Peak Ratio</th>
<th>Daily Total Occupant Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.87</td>
<td>1.65</td>
</tr>
<tr>
<td>1</td>
<td>1.55</td>
<td>1.42</td>
</tr>
<tr>
<td>2</td>
<td>2.31</td>
<td>1.83</td>
</tr>
</tbody>
</table>

- **Typical Weekly Occupancy Pattern Portrait**

Similar with daily occupancy pattern portrait, there’s also a weekly occupancy pattern portrait to demonstrate the general occupancy features of all 16 buildings.

The portraits of weekly occupancy patterns for railway stations suggest that there is always a peak on Fridays. For airports, the traffic on weekdays are significantly greater than that on weekends.

The features of the commercial complexes vary from the buildings. However, generally speaking, more occupants go to commercial complexes on weekends. This may assist with the control of building energy systems.

![Figure 9: Typical Weekly Occupancy Pattern Portrait of 16 case buildings.](image)

**Discussion**

The above-mentioned sections introduce a new approach to acquire and analyze building occupancy data, and present a series of typical occupancy patterns. In this section, the occupancy pattern from the mobile positioning data and that from energy codes are compared and discussed.

Here we take hospital as an example. Figure 10 shows the comparison of hospital’s occupancy pattern in ASHRAE 90.1 standards with that from positioning data. In the occupancy schedule from positioning data, there exists a peak during 9:00-11:00 in the morning, and the occupancy from positioning data during night time are higher than that from ASHRAE Standard. These are the major differences between two schedules.

![Figure 10: Hospital’s occupancy pattern from ASHRAE Standard and that from positioning data.](image)

In order to identify the effect of this difference on occupancy schedule, a hospital model is chosen as the example. The hospital model in located in Beijing. The energy simulation is performed by DeST-C. Figure 11 demonstrates the layout of the hospital building.

![Figure 11: The layout of the hospital model.](image)

All the settings are based on the energy codes except the occupancy schedule. One model uses code-based occupancy schedule and the other model uses positioning-data-based occupancy schedule. The following figures demonstrates the heating or cooling loads under two schedules. The cooling load of July 16th and the heating load of January 15th are plotted in Figure 12 and Figure 13. And the following table summarizes the energy statistics of the two schedules.

![Figure 12: Cooling load of ASHRAE Standard based and positioning data based occupancy schedule.](image)
occupancy features of different buildings. Finally, this research analysed the difference of final heating or cooling loads based on occupancy schedules from energy codes and from positioning data, indicating significant influence of occupancy data on energy consumption. The results suggest that mobile positioning data provides more precise information and should be considered a better approach for acquiring and applying occupancy data in both design and evaluation of energy consumption of buildings.

**Acknowledgement**

This study was supported by the China National Key R&D Program (Grant No. 2018YFC0704301) and Tsinghua University Tutor Research Fund.

**References**


