

# Anomaly Detection of Pedestrian Flow: A Machine Learning Method for Monitoring-Data of Visitors to a Building

Kentaro Kumagai<sup>1,2</sup>

<sup>1</sup>Graduate School of Management, Kyoto University  
Maskawa Bld. for Res. and Ed., Kitashirakawa-Oiwake, Sakyo, Kyoto, 606-8502, Japan  
kumagai.kentaro.3e@kyoto-u.jp

<sup>2</sup>Disaster Prevention Research Institute, Kyoto University

**Abstract** - Many public facilities such as community halls and gymnasiums are supposed to be evacuation sites when disasters occur. From the viewpoint of managing such facilities, it is necessary to monitor the usage and to respond immediately when an anomaly occurs. In this study, an integrated system of IoT sensors and machine learning for anomaly detection of pedestrian flow was proposed for buildings that are expected to be used as emergency evacuation sites in the event of a disaster. For trial practice of the system, infrared sensors were installed in a research building of a university, and data of visitors to the fourth floor of the building was collected as a time series data of pedestrian flow. As a result, it was shown that anomalies of pedestrian flow at an arbitrary time of a day with an occurrence probability of 5 % or less can be detected properly using the data collected.

**Keywords:** anomaly detection, pedestrian flow, integrated system, IoT sensor, machine learning

## 1. Background

Many public facilities such as community halls and gymnasiums are supposed to be evacuation sites when disasters occur. From the viewpoint of managing such facilities, it is necessary to monitor the usage and to respond immediately when an anomaly occurs. However, it is inefficient for humans to monitor the facilities for 24 hours. In this study, we investigate an integrated system of IoT sensors and machine learning for anomaly detection of pedestrian flow. The system monitors pedestrian flow by using IoT sensors, transmits monitoring data to a digital storage on the web in real time, detects anomalies of pedestrian flow automatically, and supports human judgment.

## 2. Integrated System of Anomaly Detection of Pedestrian Flow

Fig. 1 shows the configuration of the integrated system of anomaly detection as a conceptual diagram. IoT sensors are installed at an entrance of a building to monitor pedestrian flow. When the sensor detects visitors who enter to the building, an occurrence time of an event is recorded and the data of pedestrian flow is sent automatically in real time to a digital storage on the web. Subsequently, a computer with machine learning program analyses the data (Fig. 2): The computer performs a classification of the time-series data of the past data up to the previous day, into some classes based on

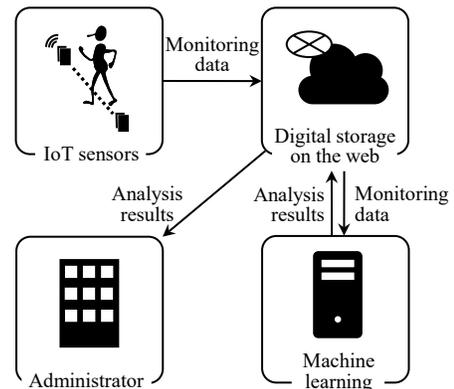


Fig. 1: Configuration of Integrated System.

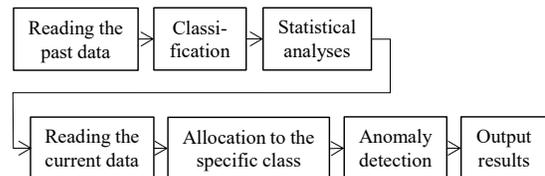


Fig. 2: Flow diagram of Machine learning.

characteristics of the data. And the computer conducts statistical analyses on the data of each class. After the process is finished, the computer allocates the current day's data to the specific class based on the similarity, and performs anomaly detection on the data. Here, "the current day's data" means a set of time-series data of pedestrian flow from the start of the day up to the current time. As the time passes, a data set of "the current day's data" is updated, and the computer repeats the above process iteratively, based on the latest information. At each time step during the day, the computer automatically outputs results and informs them to an administrator when abnormality of pedestrian flow is detected, to support human judgement.

### 2.1. Trial Practice of the system

The author conducted trial practice of the system as follows: A pair of infrared sensors was installed

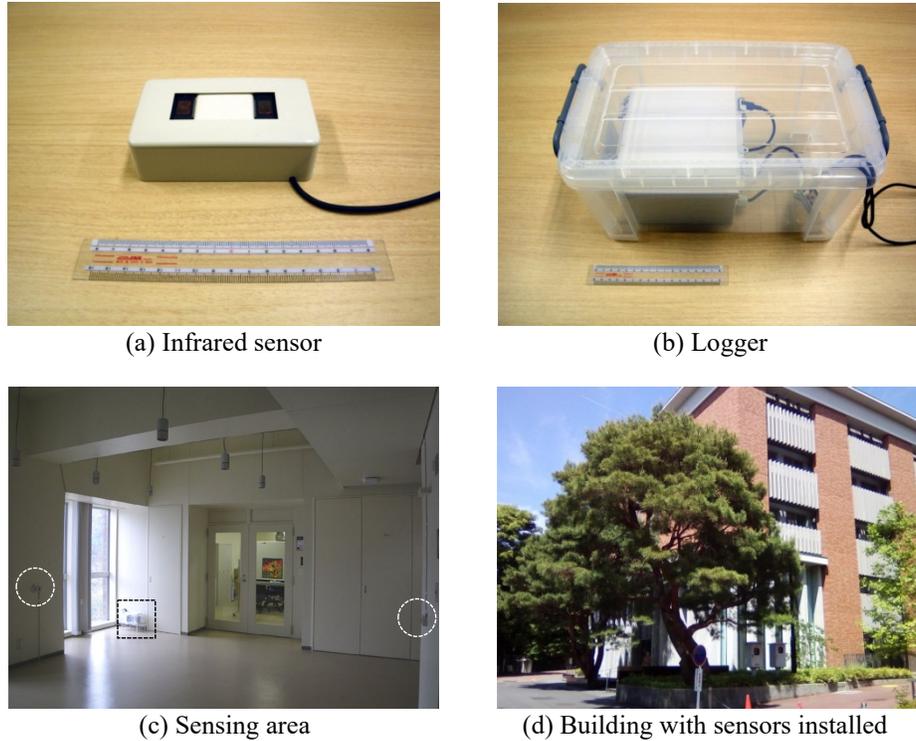


Fig. 2: Infrared sensor, Logger and Building with sensors installed.

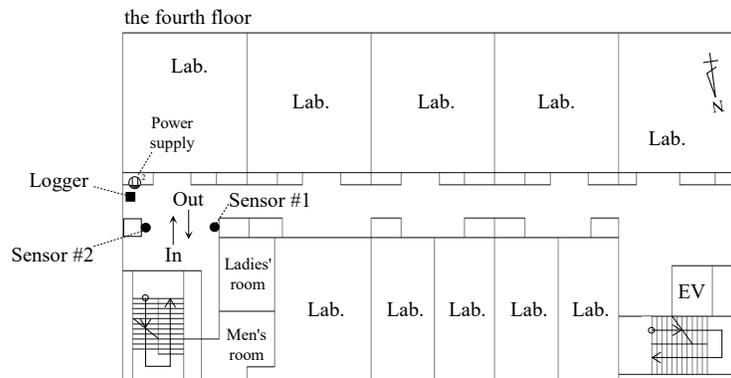


Fig. 3: Plan view of the sensing area in the building

at an entrance of the fourth floor of a research building in Kyoto University. The sensors and a logger (Fig. 2(a), (b)) were products of Scanmatic Marketing Co., Ltd., 'i-trend mini' package. The sensors were attached on the wall at the height of 1 m from the floor, and the sensors were possible to detect the visitor walking through the space between the two sensors. Because a pair of sensors included two irradiation units, the sensor was possible to detect walking direction of the visitor. The minimum resolution in time of this sensor was approximately 0.3 s per event. Each sensor was connected to the logger by cables. Fig. 2(c) shows the sensing area, two white dotted circles on the photo indicate the sensors respectively, and the black dotted square indicates the logger. Fig. 3 is a plan view of the sensing area. The logger saves an occurrence time into a SD memory card, when the sensor detects visitor in 'IN' direction of the figure. The electricity consumed by the sensors and the logger was very low.

Data collected was classified manually into two classes, such as data of weekdays and data for school holidays. A FORTRAN program was written for calculation of the statistical values from data for each class. After the mean value  $\mu$  and standard deviation  $\sigma$  were calculated for each class by using the past data

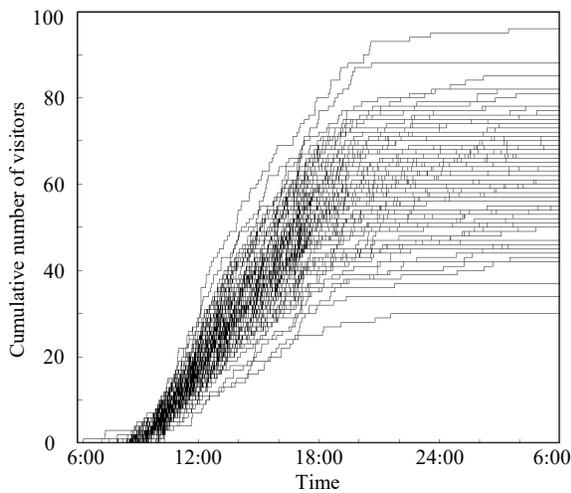
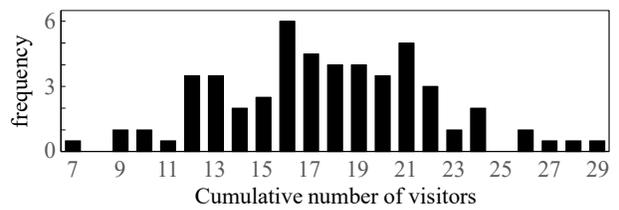
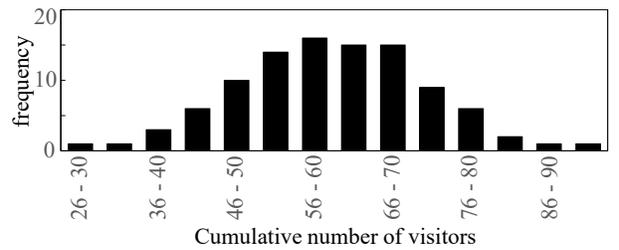


Fig. 4: Time-series of the cumulative number of visitors on weekday.



(1) At 12:00



(2) At 24:00

Fig. 5: Histogram of the cumulative number of visitors.

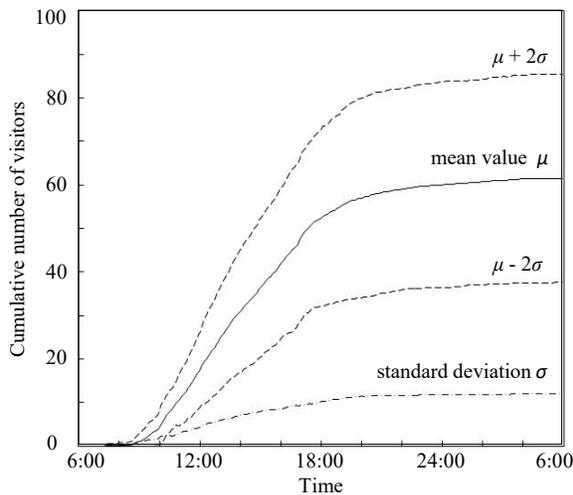


Fig. 6: Mean value  $\mu$  and standard deviation  $\sigma$ .

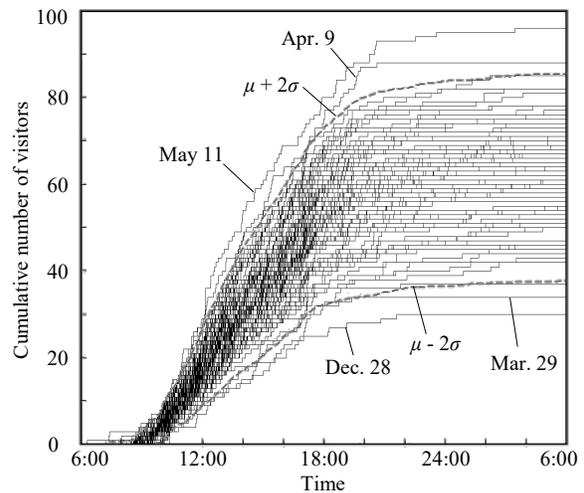


Fig. 7: Superimposing  $\mu+2\sigma/\mu-2\sigma$  on Fig. 4.

up to the previous day, the program checked "the current day's data" of pedestrian flow at every five minutes, in view of that the value was in the range of  $\mu - 2\sigma$  to  $\mu + 2\sigma$  or not.

## 2.2. Results

Based on the data collected for 152 days, from December 11, 2017 to May 11, 2018, the cumulative number of visitors per day was seemed to be different between weekdays and school holidays. On weekdays, the cumulative number of visitors was in the range of 30 to 95 persons per day. In contrast, that on school holidays was 20 or less. Therefore, it seems to be reasonable to divide the acquired data into two classes, such as data of weekdays and data of school holidays. Here, it should be noted that the word 'a day' means twenty four hours from 6:00 am to 5:59 am of the next day, in this section and the following sections.

Fig. 4 draws the time series of the cumulative number of visitors of weekdays during the above period. The numbers of weekdays were a hundred. The horizontal axis means time, and the vertical axis means the cumulative number of the visitors. It was also possible to draw the distribution of the cumulative number of visitors at an arbitrary time of a day. As an example, Fig. 5(1) and (2) are the histograms of cumulative number of visitors at 12:00 and at 24:00 on weekdays, respectively. In Fig. 5(2), the values were distributed among 30 to 95. Although a strict validation has not been done yet, it seems to be possible to assume that the number of visitors at 24:00 follows a normal distribution. The mean value  $\mu$  and the standard deviation  $\sigma$  were 61.5 and 12.0, respectively. Fig. 6 shows time-series of  $\mu$  and  $\sigma$  in a solid line and a dot-dash line, respectively, calculated at every five minutes from 6 am. In the figure, the other two broken-lines showed values of  $\mu + 2\sigma$  and  $\mu - 2\sigma$ , respectively. Assuming that the number of visitors at an arbitrary time of a day follows a normal distribution, if a value exceeding  $\mu + 2\sigma$  or a value below  $\mu - 2\sigma$  is observed at an arbitrary time of a day in pedestrian flow, it can be judged in real time to occur an anomaly with 5 % or less probability. In Fig. 7, time zones such as almost whole day of December 28, March 29, April 9 and May 11, respectively, should be detected as anomalies.

## 2.3. Discussions

*Dynamic algorithm for automatic classification:* To conduct the anomaly detection, information about a class which the current day belongs to is necessary. In the sections 2.1 and 2.2, allocation to the two classes, i.e. weekday or school holiday, was done manually. In case that the target facility is a school, school holidays are often known in advance. However, there are possibilities that irregular holidays may occur depending on facilities. For such facilities, a dynamic algorithm will be requested for automatic classification which allows a user of the system to sort automatically training-data and "the current day's data" up to the current time to an appropriate class dynamically. Since the system is intended to be applied to real-time anomaly detection of pedestrian flow, further consideration is necessary for that point.

*Periodic variation:* Fig. 8 is a power spectrum of the time-series data of the cumulative number of visitors per day for 152 days, from December 11, 2017 to May 11, 2018. The horizontal axis shows a period of time in the log scale, and the vertical axis shows a density of the power spectrum. The highest peak was at 6.91 day, and this fact means that the strongest periodic fluctuation components included in the data set was a period of approximately one week. In Section 2.2, we focused on weekdays. It was

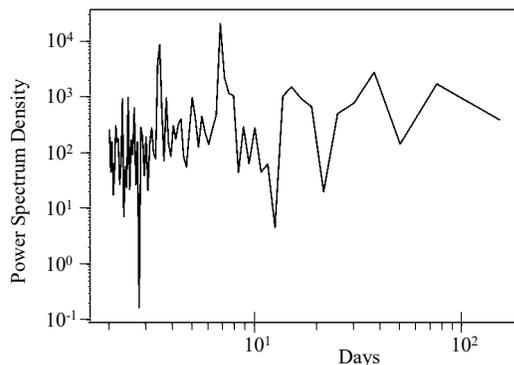


Fig. 8: Power Spectrum

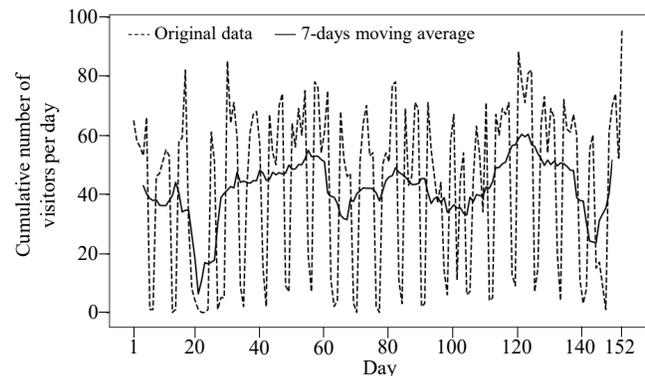


Fig. 9: 7-days moving average

consistent with the fact that the strong periodic fluctuation component of one week.

Fig. 9 shows 7-days moving average of the same data set. The horizontal axis shows number of days since measurement started, and the vertical axis shows the cumulative number of the visitors per day. The broken line and the solid line are the original value and the 7-days moving average, respectively. The broken line obviously showed periodic fluctuation in approximately one week. The solid line showed that the values of the 7-days moving average were at the range of 40 to 60 in many days, but there were four small valleys. The first valley corresponded to the holiday season of year-end of 2017 and beginning of 2018. The second and the third valleys were at the middle of February and the middle of March, respectively. These are both included in the end term of school year of Japan. And the last valley corresponded to another holiday season in May. Since data after the period used in this paper has not been acquired yet, long-term periodic fluctuation, such as monthly, quarterly and yearly, is unknown. Data acquisition is ongoing, and further consideration is necessary when the data becomes available.

### 3. Conclusion

In this study, an integrated system of IoT sensors and machine learning for anomaly detection of pedestrian flow was proposed for buildings that are expected to be used as emergency evacuation sites in the event of a disaster. For trial practice of the system, infrared sensors were installed in a research building of a university, and data of visitors to the fourth floor of the building was collected as a time series data of pedestrian flow. As a result, it was shown that anomalies of pedestrian flow at an arbitrary time of a day with an occurrence probability of 5 % or less can be detected properly using the data collected.

### Acknowledgements

The author would like to express appreciation to Prof. Tomihiro Watanabe, Dr. Masashi Ogawa and Miss Misato Hiraoka of Graduate School of Management, Kyoto University for their support of data collection by sensors at the Maskawa Building for Research and Education of Kyoto University.

### References

- [1] T. Ide, Introduction of Anomaly Detection by Machine Learning - Practical Guide by R, Corona Publishing Co.,Ltd., ISBN 9784339024913, 2015.
- [2] T. Ide and M. Sugiyama, Anomaly Detection and Change Detection, Machine Learning Professional Series, Kodansha Ltd., ISBN 9784061529083, 2015.
- [3] M. Jin, Data Science by R - from the Foundation of Data Analysis to the Latest Method, Morikita Publishing Co., Ltd., ISBN 9784627096028, 2017.