An analysis and evaluation of spatiotemporal behaviour of people through Wi-Fi sensing data

Guanghui Zhou

Mechanical and Civil Engineering Division

Graduate School of Engineering

Gifu University

2022

Ask not that events should happen as you will,
but let your will be that events should happen as they do,
and you shall have peace.

— Epictetus

Acknowledgement

In September 2017 I started my life in Japan and after half a year as a research student, I started my Ph.D. life in the Mobility analytics laboratory at Gifu University. Ph.D. life refers to a new and challenging adventure. Along this journey, I learned that nothing is as one's expects and there are always problems along the road to success. What you can do is never give up until complete. I believe the experience learned during the Ph.D. life stage is invaluable for me. It gives me the confidence to deal with problems that even I never met before. Many colleagues, friends, and family members have helped me a lot during this amazing journey, particularly in another country. I am immensely grateful to all of them.

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Fumitaka Kurauchi. Without his guidance, advice, encouragement, patience, and impressive knowledge, this thesis would not have been possible. His vision and enthusiasm have motivated me to grow as a research scientist and enriched my way of thinking. He provides many opportunities to make presentations at various conferences and seminars and encourages me to communicate with others. I am thankful to him for all the support.

Besides my supervisor, I would like to express my gratitude to those who reviewed the manuscript and made useful comments and suggestions as committee members: Prof. Toshihiko Miyagi, Prof. Akiyoshi Takagi, Prof. Yoshifumi Demura, and Prof. Jiangqian Ying. Your comments help me to improve the quality of my thesis from various perspectives. In fact, not only during the defense, but I also got valuable comments from them on other occasions such as Miyagi Meeting, the joint presentation between 3 labs (Kurauchi Lab, Takagi Lab, and Demura Lab), and the Gursps.

I am very grateful to those who also made valuable suggestions about this work, such as during the joint seminar between 6 Universities (Kyoto University, Ritsumeikan University, Nagoya University, Miyazaki University, Hokkaido University, Gifu University) and the CART project meeting. They are Prof. Nobuhiro Uno, Associate Prof. Jan-Dirk Schmöcker, Prof. Yasuhiro Shiomi, Associate Prof. Toshiyuki Nakamura, Associate Prof. Hiroshi Shimamoto, Associate Prof. Satoshi Sugiura, Prof. Junji Nishida. Thank you for making such enjoyable opportunities for communication and especially for the 6-University seminar because there is BBQ after the presentation, and we can chat and make friends while drinking beer. As for Associate Prof. Satoshi Sugiura, who worked at Gifu University before, I still remember the first day I came to the lab, you invited me to your office when I was confusing in the hallway…how time flies!

I am also grateful to Dr Takahiro tsubota at Ehime University. I truly appreciate your valuable comments and the discussions we have had.

My profound thanks are extended to the company JRISS for providing Wi-Fi packet sensor data; Mr Nishida and Mr Asao as well as their colleagues for the development of the Wi-Fi packet sensor and providing the data.

I greatly appreciate the help provided by those who helped with the instalment of the sensors and manual counting survey. That was fundamental to completing my work.

I would also like to thank Mr Komori, Mr Sugiyama, and Mr Ban at Teikoku company for kindly providing me with the internship opportunity.

My gratitude is also extended to the laboratory secretary, Maki Aso, for your kind assistance with the procedure materials and other daily affairs.

I greatly appreciate the kind support I received from all the present and former laboratory members for their concerns and friendships during my study and life. It is a very great experience in the Mobility analytics laboratory, at Gifu University. I like the atmosphere and pleasant environment here.

Special thanks to my wife, Dr Zhang, for your love, encouragement, patience, and our upcoming baby.

My most deep thanks go to my parents who are the base of all my achievements. To my older sister, thanks for your support and encouragement.

For anyone whom I didn’t explicitly mention, I would like to express my gratitude for the time they shared with me.

Guanghui Zhou
October 2022

Preface

Parts of this dissertation have been presented in conferences or else submitted for review are as follows.

Conference presentations and publications

I. **Guanghui Zhou,** Fumitaka Kurauchi, Shuhei Myoko. (2019). On the characteristics of Wi-Fi packet sensors for traffic and pedestrian flow observation. 60th Japan Infrastructure Planning Conference (Autumn Meeting), Toyama, Japan. (Oral presentation, **chapter 3**)

II. **Guanghui Zhou,** Fumitaka Kurauchi, Shuhei Myoko. (2019). On the characteristics of Wi-Fi packet sensors for traffic and pedestrian flow observation. 24th International Conference of Hong Kong Society for Transportation Studies, Hong Kong. (Poster presentation)

III. **Zhou, G.**, Kurauchi, F., Ito, S., & Du, R. (2022). Identifying golden routes in tourist areas based on AMP collectors. Asian Transport Studies, 8, 100052. (Presented in 14th EASTS international conference, **Chapter 6**)

IV. **Guanghui Zhou**, Fumitaka Kurauchi, Shuhei Myoko. (2021) Analysis on the characteristics of AMP based traffic and pedestrian flow observing system. 27th ITS World Congress, Hamburg, Germany, pp.510-521. (Oral presentation, **Chapter 3**)

Submitted

I. **Zhou, G.**, Kurauchi, F. (2022). Pedestrian level of service evaluation based on Wi-Fi packet sensor data. 66th Japan Infrastructure Planning Conference (Autumn Meeting), Okinawa, Japan. (**Chapters 4 and 5**)

Abstract

With the growth of world-wide population and urbanisation, crowd analysis has received lots of attention from social and technical disciplines and has been an actively growing research area. Meanwhile, the advances in means of transportation have facilitated the gathering of people, many people have become indulged in travelling and the tourism industry has also expanded significantly across the globe. Tourism brings economic benefit and contributes to the employment and development in destination regions. However, the concentration of tourist flows to specific areas may lead to overcrowding of destinations and generate negative impacts such as alienated local residents, degraded tourist experiences, overloaded infrastructure, harm to the natural and cultural heritage, environmental pollution and transportation congestion. To overcome these issues and ensure public safety and make areas more convenient, understanding human spatiotemporal movement features and crowd phenomenon is of great importance and also plays a key role in a variety of application domains such as store location, developing crowd management strategies in public events as well as planning and designing of public spaces, and providing guidelines for navigation in large buildings such as train stations, airports, stadiums and theatres and so on.

Tourism is one of the pillar industries of Japanese economic development. As one of the most famous tourist attractions of Japan, Kyoto is also faced issues that come from overtourism such as queuing at a bus station and transportation congestion. How to reasonably guide and manage tourists, provide a comfortable travel experience to visitors and reduce negative impacts on residents, have become urgent problems. Therefore, it is increasingly important to analyse and understand the movement behaviour and features of visitors to alleviate congestion in tourist areas and improve services offered to visitors and citizens.

Pedestrian Level of Service (PLoS) is one of the best criteria to characterise the performance of a given road in terms of travellers' perspective and is often described as the comfort level that is experienced by the pedestrian. The Highway Capacity Manual (HCM) provided certain guidelines for calculating pedestrian levels. In tourism research, Sequential Pattern Mining (SPM) has been extensively used by researchers to understand the destination visiting behaviour of tourists for efficient destination management and attraction marketing. Therefore, the PLoS and SPM are used as tools to study the crowding behaviours in this research. The dynamic nature of visitor flows and destination visiting trajectory information are key components of these two tools. The most common people data acquisition methods include manual count surveys and video surveillance. However, high labour costs and difficulty to

acquire long term data are always a problem for manual count surveys. Video camera-based data collection is also dependent on weather conditions, illumination changes, and limited viewing angles. Moreover, it is difficult to track the crowds because the camera-based data collection requires image processing to identify and track people, while it is common for blocking each other in dense crowds. On the other hand, thanks to the development of information and communication technology, smartphones have become more and more popular in the past decade. One smartphone has many kinds of sensors and functions which can be used to record many kinds of information about the user. The development of radio frequency technology provides more possibilities for using this information. For example, we can collect pedestrian movement data by tracking the location of their smart devices since they carry smartphones almost anytime and anywhere. Wi-Fi packet sensor data has been applied to analyse pedestrian behaviour under different scenarios such as the customers within a shopping mall, and passengers in a transit station. This data acquisition method requires less effort and fewer resources yet produces a larger volume of data than traditional counting and surveying. However, it is still not well developed in the current situation, especially limited research has been done in applying this technology to an outdoor context and analyses of a specific group (tourist). There are still challenges facing it such as how the environmental and installation conditions influence the observation result of the Wi-Fi packet sensor, and the correlation between the sensor observation and real pedestrian count is not clear. Therefore, this research carried out an experiment to collect data from a famous tourist area to extend the potential application of anonymous Wi-Fi sensing technology in various contexts. Moreover, other experiments were also carried out in the laboratory and campus to collect data as fundamental analysis of MAC address data for tracking people.

Overall, this research aimed at unlocking and making available hidden data to analyse crowd behaviour. In this research, the Wi-Fi packet sensor was used to collect the data of smart device users. This is a passive data collecting method, without the cooperation of users, without installing any apps and without infringing on privacy. Specifically, the main contributions of this thesis can be summarized as:

Chapter 3 explored the factors that influence the observation of the Wi-Fi packet sensor and gave a preliminary proof that the Wi-Fi packet sensor can be used to analyse human movement.

Chapter 4 quantified the influencing factors and developed a method to estimate real pedestrian count based on Wi-Fi packet sensor data and manual count data collected at Higashiyama area, Kyoto. The result of this chapter is the foundation of chapter 5.

Chapter 5 used the concept of PLoS (Pedestrian level of service) to evaluate the crowding level in a tourist area. In this chapter, it is possible to estimate real pedestrian flow with the parameters estimated in chapter 4, based on which the PLoS can be decided. Possible solutions to balance people in time and space are also discussed.

Chapter 6 proved that the characteristics of smart device users can be identified through clustering analysis even data from the Wi-Fi packet sensor is anonymous. It also extracted the frequently used routes by tourists in Higashiyama area. These analytical methods may be applicable to other tourist destinations and pedestrian flow studies, such as passenger flow inside a transfer station. This technology can also help monitor the pedestrian travel changes before and after the COVID-19 pandemic. For example, it can observe the volume and stay time of pedestrians at a public place like a mall or restaurant and can also observe the use of public transportation and cross-city mobility.

This research showed it is possible to analyse crowd behaviour utilising Wi-Fi sensing technology, especially in terms of tourist behaviour. This can help local government and tourism destination managers make strategies for tourist control and tourism management. The analyses of this research provide evidence and insights for the further application of this new data source.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background and research motivation

1.1.1 Crowd analysis

The opportunity for jobs, convenience and prosperity, among other factors, pulls people to cities. According to the statistical data of the United Nations, more than half of the world's population now lives in cities and it has been projected that 68% of the world population will live in urban areas by 2050. The high population density can cause problems such as lack of open space, pollution (land, pollution, water pollution and air pollution), overcrowded, and traffic congestion. Effective city planning and management by national and local authorities will be essential in addressing these and other issues as the world's urban areas swell.

With the growth of world-wide population and urbanisation, crowd analysis has received lots of attention from social and technical disciplines and has been an actively growing research area (Keith Still, 2000; Lamba and Nain, 2017; Kaiser et al., 2018; Bendali-Braham et al., 2021). Understanding human spatiotemporal movement features and crowd phenomenon is of great interest and critical in a variety of application domains such as store location, developing crowd management strategies in public events as well as planning and designing of public spaces, and providing guidelines for navigation in large buildings such as train stations, airports, stadiums and theatres. The goals include ensuring public safety and making areas more convenient (Zhan et al., 2008; Timmermans, 2009). Meanwhile, many people have become indulged in travelling and the tourism industry has also expanded significantly across the globe. According to the statistics of the United Nations World Tourism Organisation (UNWTO), there were 1.5 billion international tourist arrivals in 2019 and it was expected that the international flow of tourists will reach 1.8 billion by 2030, globally. Japan is also one of the most attractive countries for tourism. Based on the data of the Japan National Tourism Organization (JNTO), Figure 1.1 shows the number of international visitors to Japan from 2011 to 2020. It can be recognised that the number of international visitors to Japan has increased more than five times in the recent decade. With the outbreak of Covid-19, this trend has been abruptly reversed now. Due to the restriction on international travel, the tourism industry has been hit extremely hard and the tourism economy also experienced a huge decline. Even if the problem of overcrowding could be released, it is unrealistic and not a long-term policy to restrict international travel. The ongoing pandemic could provide us with an opportunity to rethink tourism as a whole by shifting more decisively from overtourism to sustainable tourism.

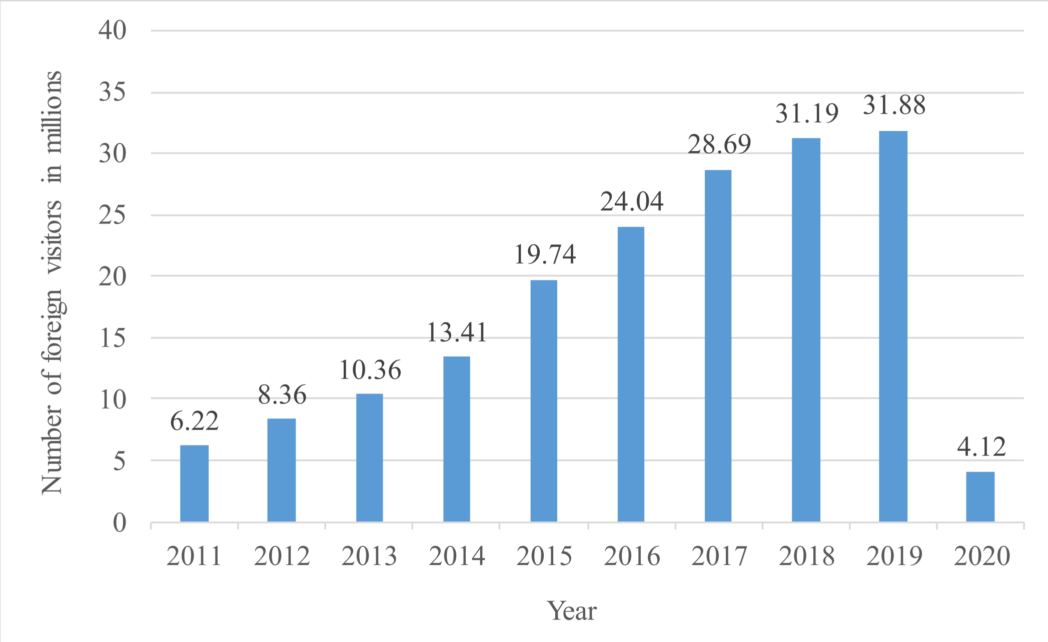

Figure 1.1 Number of international visitors to Japan from 2011 to 2020

Tourism brings economic growth and contributes to the employment and development in destination regions. However, the concentration of tourist flows to specific areas may lead to overcrowding of destinations and generate negative impacts such as alienated local residents, degraded tourist experiences, overloaded infrastructure, harm to the natural and cultural heritage, environmental pollution and transportation congestion. (McKinsey & Company, & World Travel & Tourism Council, 2017; Peeters et al., 2018). The problems caused by the ongoing growth in tourism have created the issue of overtourism and it was found that overtourism has emerged rapidly as a concept and crowding/overtourism has become an important issue for residents as well as tourists in several destinations, including New York City, Amsterdam, Reykjavik, the Isle of Skye, the Koh Phi Phi Islands of Thailand and the Palawan archipelagic province of the Philippines (Peeters et al., 2018; Capocchi et al., 2019; Pechlaner et al., 2019). The World Tourism Organization (UNWTO) defines overtourism as an impact of tourism on a destination, or parts thereof, that excessively influences the perceived quality of life of its citizens and/or the quality of visitors' experience in a negative way (UNWTO, 2018). This report also clearly states that the tourism must be developed and managed sustainably for both visitors and local communities, as tourism is an opportunity for communities and their people to share its benefits.

Tourism is one of the pillar industries of Japanese economic development. As one of the most famous tourist attractions of Japan, Kyoto attracted more than 53 million tourists (more than 7 million coming from overseas) based on the statistical data of 2017.

Figure 1.2 Tourists in the streets around Kiyomizu temple, in Kyoto, in the pre-pandemic era.
Photo: Shutterstock

Tourists have not only brought economic benefit to Kyoto city but have also caused many problems. For example, an increase in tourists has caused severely crowded conditions inside buses, which brings discomfort to residents and too many pedestrians can also easily cause traffic congestion. Kyoto has fallen into an overtourism situation (Ken Victor Leonard, 2020;Satake et al., 2019; Lee, 2021). While the Japanese Government continues to develop tourism to promote the economy, how to reasonably guide and manage tourists, provide a comfortable travel experience to visitors and reduce negative impacts on residents, have become urgent problems. Therefore, it is increasingly important to analyse and understand movement behaviour and features of visitors to alleviate congestion in tourist areas and improve services offered to visitors and citizens. UNWTO (2004) claims that sustainable tourism industry depends on effective management of tourism flows at and through destinations and sites, giving the visitor time and opportunity to appreciate and enjoy the local culture and the values of the places being visited and to acquire local goods and services. Oklevik et al. (2019) analyses overtourism concerning the crowding effect, observing how crowding as an issue for destinations has been a recurrent topic in tourism research since the early 1970s (Turner and Ash, 1975; Ward and Berno, 2011). According to the research of (McKinsey & Company, &

World Travel & Tourism Council, 2017), destinations can mitigate overcrowding by adopting the right mix of tactics which include smoothening visitors over time and spreading visitors across sites, adjusting pricing to balance supply and demand, regulating accommodation supply and limiting access and activities.

1.1.2 New technologies for collecting pedestrian data

If the movement behaviour of pedestrians can be understood spatially and temporally, it will be helpful to take measures to control the pedestrian flow to alleviate crowding and enhance the pedestrian experience. According to UNWTO (2004), improving visitor movement patterns around sites can help to manage tourism better to avoid congestion at tourist sites. For example, ensure that the site is regularly monitored, especially in peak periods, to confirm that the movement pathways are working efficiently and ensure that the visitor movement patterns are continuous or in a one-way circulation system, to avoid returning visitors competing with those walking to the attraction are some efficient measurements to improve pedestrian movement. As a way to incorporate the density experience of a pedestrian into infrastructure design, Fruin (1971) introduced the Level-of-Service concept for pedestrians. Pedestrian Level of Service (PLoS) is one of the best criteria to characterise the performance of a given road in terms of travellers' perspective. The Highway Capacity Manual (HCM) used this concept as a measure to describe operational conditions of pedestrian traffic and provided certain guidelines for calculating pedestrian level (HCM, 2000). The Level of Service is often described as the comfort level that is experienced by the pedestrian (Bloomberg and Burden, 2006). In tourism research, Sequential Pattern Mining (SPM) has been extensively used by researchers to understand the destination visiting behaviour of tourists for efficient destination management and attraction marketing (Xia et al., 2005; Lew and McKercher, 2006; Orellana et al., 2012; Bermingham and Lee, 2014; Bin et al., 2019; Park et al., 2020; Abucejo and Cuizon, 2021). Therefore, the PLoS and SPM will be used as tools to study the crowding behaviours in this research. The dynamic nature of visitor flows and destination visiting trajectory information are key components of these two tools. The most common people data acquisition methods include manual count surveys and video surveillance. However, high labour costs and difficulty to acquire long term data are always a problem for manual count surveys. Video camera based data collection is also dependent on weather conditions, illumination changes, limited viewing angles (Liebig and Kemloh Wagoum, 2012) and other factors also can result in lower recognition rates such as complex background, shadows and abrupt motion (Gawande et al., 2020). Another major shortcoming of video-based human data collection is that it is difficult to unambiguously distinguish between people in a crowd because of constant interactions and blocking of each other. This severely restricted its use as a tracking method for the analysis of pedestrian behaviour in space utilisation (Dee and Velastin, 2008). On the other hand, with the rapid development of computer science and Internet techniques, massive-scale data are generated, recorded, stored and accumulated, forming the big data and opening a new age, various big data sources have been applied to enrich and promote tourism research (Kambatla et al., 2014; Li et al., 2018). Technological advancements have created a variety of data acquisition methods that require less effort and fewer resources yet produce a larger volume of data than traditional counting and surveying, as is the case with new technology-based data available through mobile devices. Automatic counting techniques are the most promising strategy for enhancing the amount and quality of such data. The number of smartphone users worldwide today surpasses three billion and is forecasted to grow further by several hundred million in the next few years ("Number of smartphone users worldwide from 2016 to 2021| Statistic," n.d.). Besides, wireless internet access has become a standard feature of smartphones and each smartphone has a media access control (MAC) address that is unique to each device. The unique identifiers can be matched over space and time; therefore, the data are ideal for tracking devices and can be the potential for understanding pedestrian movement behaviour. This information means that Wi-Fi probe request data sources are becoming more and more massive, and the Wi-Fi probe detection-based data collection method called anonymous MAC address packet (AMP) sensing is becoming increasingly useful. Wi-Fi packet sensor data has been applied to analyse pedestrian behaviour under different scenarios such as the customers within a shopping mall (Fukuzaki et al., 2015), passengers in a transit station (Schauer et al., 2014; Hwang et al., 2019) or the students' movement and occupancy of a campus (Kalogianni et al., 2015; Andión et al., 2018) and in dense urban environments(Traunmueller et al., 2018). The Wi-Fi packet sensor data can also be used to support smart cities concept (Kyritsis, 2017).

Although multiple studies make use of Wi-Fi sensing technologies to analyse pedestrian behaviour, the potential of using this new emerging data source still needs to be explored and broadened in various contexts and perspectives. To this end, this research focuses on analysing and understanding pedestrian behaviour through the passive Wi-Fi sensing data and exploring the applications in tourism management.

1.2 **Research questions and objectives**

Addressing the aforementioned main purpose, there are still challenges in using MAC address data as a tracking technology for the monitoring of pedestrians. Our main research question can be broken into several smaller ones. Firstly, in order to collect efficient data and as much data as possible, it needs to identify the influencing factors of the observation result of this technology. This brings us to the first research question:

Question 1: What factors affect the data collection of the Wi-Fi packet sensor and whether it is suitable for collecting pedestrian data utilizing the Wi-Fi packet sensor?

Objective of chapter 3: This part is a fundamental analysis of MAC address data collection. To answer the question 1, indoor and outdoor experiments were carried out with 5 sensors.

Question 2: How do these factors influence the observation result of the sensor and how the pedestrian count can be estimated?

Objective of chapter 4: This part quantified the influencing factors on the sensor data collection process and built the correlation between the Wi-Fi packet sensor observations and real pedestrian flow volume. The work of this part is the foundation of chapter 5.

Question 3: Whether it is possible to evaluate the crowding level of visitors of a street?

Objective of chapter 5: This part attempts to monitor the crowding level with the concept of PLoS. Based on the installation conditions of the Wi-Fi packet sensors and sidewalk width, it is possible to monitor the pedestrian flow performance. This information can help to manage the pedestrian flow in the tourist area.

Question 4: Whether it is possible to analyse the behaviour of a specific group (tourist) with this anonymised data?

Objective of chapter 6: To gain an insight in the capabilities of this technology, this part attempts to apply this data source to the tourism analysis. Firstly, the properties of the resulting data were clustered to identify the different types of smart device users. Then the frequently used routes by tourists were extracted.

1.3 **Dissertation framework and methodology**

1.3.1 Dissertation framework

This dissertation consists of seven chapters. After the brief introduction that illustrated the background and motivation, research objective, and the structure of this dissertation in this chapter, seven chapters will be presented as follows:

Chapter 2 provided a review of the literature relevant to different pedestrian data collection methods. In this chapter, the advantages of the Wi-Fi packet sensor-based pedestrian data collection system were presented through comparison with other data collection methods such as the manual counting method and camera-based method. A detail review of literature that focuses on the specific investigated topic can be found in each related chapter. The sensor specifications and observation data characteristics are also described in this Chapter.

Chapter 3 explored the factors affecting observation results of Wi-Fi packet sensors and how accurately pedestrian and traffic flow volume can be estimated through one sensor method. The data used in this chapter were collected from experiments in the laboratory and at Gifu University Campus.

Chapter 4 developed a method to estimate real pedestrian volume based on Wi-Fi packet sensor data and manual count survey data collected in the Higashiyama area, Kyoto. The aim of this chapter is to quantify the influence factors of observation results to estimate the pedestrian volume in the Higashiyama area which will be used to evaluate the PLoS in chapter 5.

Chapter 5 explored the possibility of evaluating pedestrian flow performance based on Wi-Fi packet sensor data. To know the efficiency of roadways in aspect to accommodate pedestrian travel or evaluate whether the tourists can have a comfortable walking experience, PLoS needs to be assessed. This assessment helps for the tourist management and improvement of service.

Chapter 6 analysed the travelling behaviour of tourists in the Higashiyama area using digital footprint data collected by 20 Wi-Fi packet sensors. The clustering analysis was performed firstly to identify the trajectory of tourists from the anonymous data. Then sequential pattern mining was used to extract the frequent sequence of destinations visited by tourists.

Chapter 7 summarised the findings and provided the main conclusions of this dissertation as well as possible further research directions.

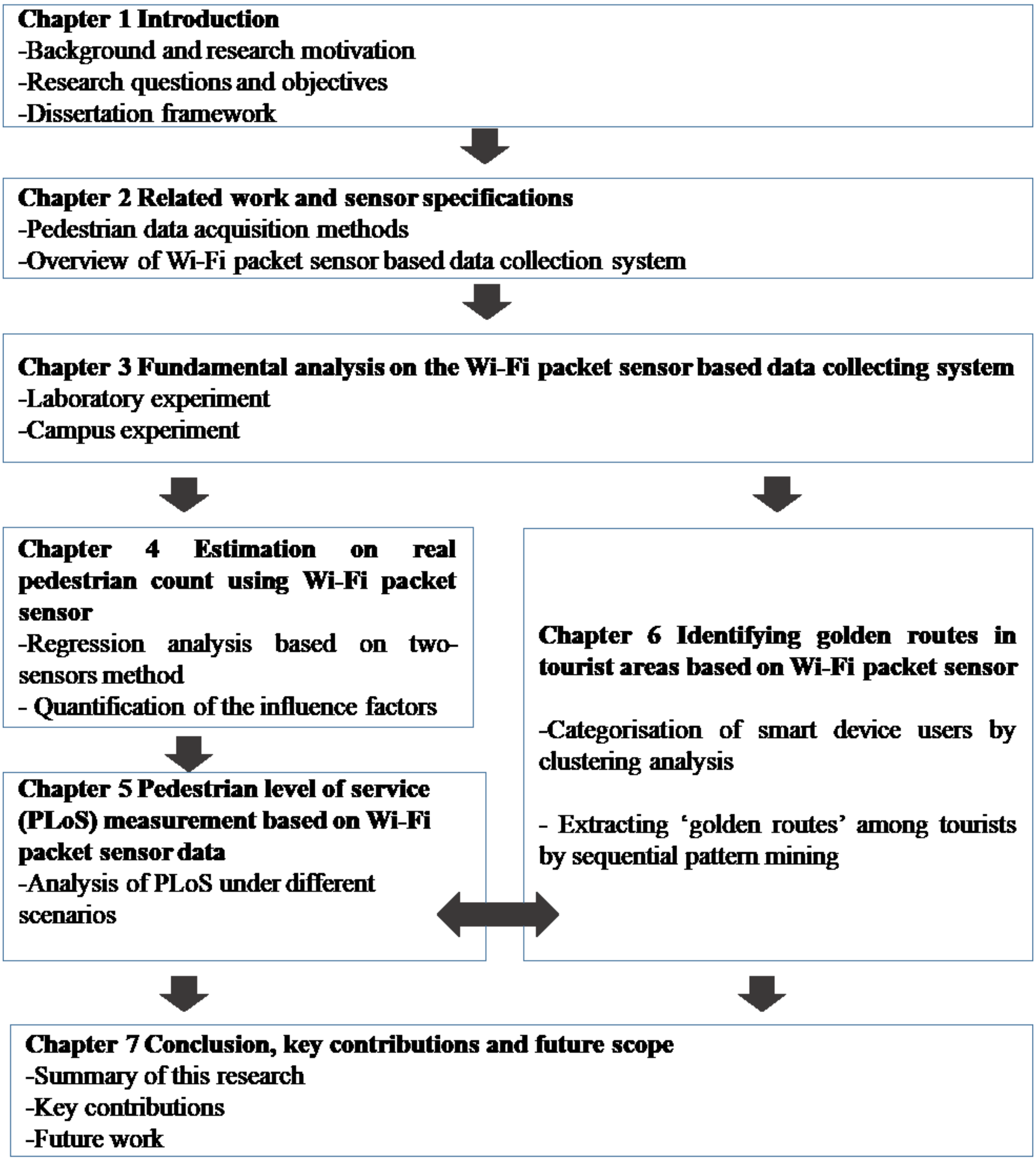

Figure 1.3 Outline of thesis

1.3.2 Methodology

In this study, a methodology to use Wi-Fi sensing data for understanding the spatiotemporal behaviour of people is proposed. Firstly, a fundamental experiment was carried out to explore the factors that influence the observation of Wi-Fi packet sensors. A trial regression and clustering analysis were also performed using the collected data. Afterwards, the influencing factors were quantified and a model to estimate real pedestrian count was built. The research methodology also includes the evaluation of pedestrian crowding level and clustering analysis to identify the types of smart device users and sequential pattern mining to identify the frequently used routes of tourists. The complete research methodology is visualised in Figure 1.4.

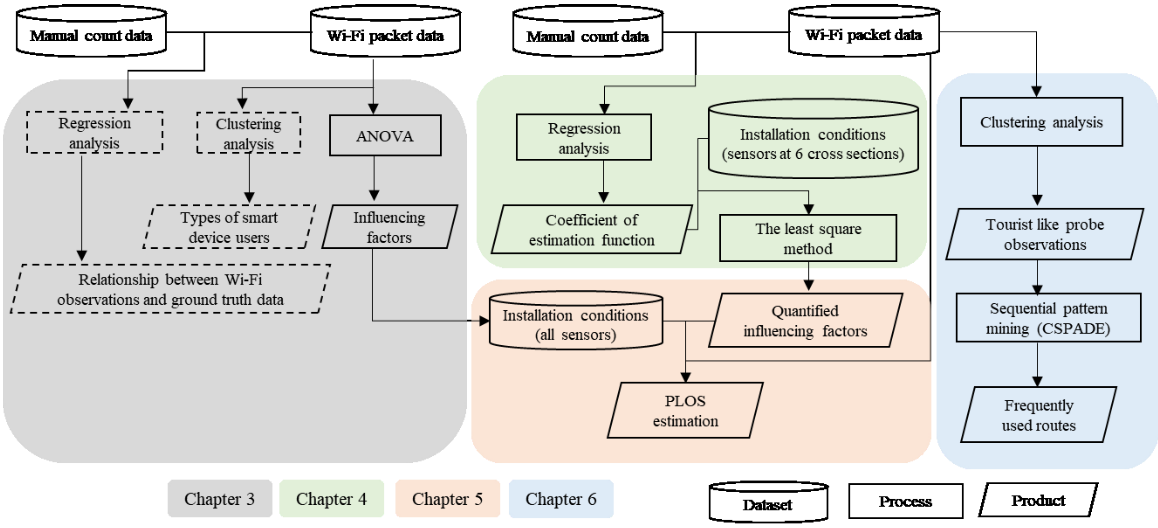

Figure 1.4 Visualisation of the research methodology

References

Abucejo, M.I., Cuizon, J., 2021. Sequential Pattern Mining of Tourist Spatiotemporal Movement. Recoletos Multidisciplinary Research Journal 9, 69–78. https://doi.org/10.32871/rmrj2109.01.07

Andión, J., Navarro, J.M., López, G., Álvarez-Campana, M., Dueñas, J.C., 2018. Smart Behavioral Analytics over a Low-Cost IoT Wi-Fi Tracking Real Deployment. Wireless Communications and Mobile Computing 1–24. https://doi.org/10.1155/2018/3136471

Bendali-Braham, M., Weber, J., Forestier, G., Idoumghar, L., Muller, P.-A., 2021. Recent trends in crowd analysis: A review. Machine Learning with Applications 4. https://doi.org/10.1016/j.mlwa.2021.100023

Bermingham, L., Lee, I., 2014. Spatio-temporal Sequential Pattern Mining for Tourism Sciences. Procedia Computer Science, 2014 International Conference on Computational Science 29, 379–389. https://doi.org/10.1016/j.procs.2014.05.034

Bin, C., Gu, T., Sun, Y., Chang, L., 2019. A personalized POI route recommendation system based on heterogeneous tourism data and sequential pattern mining. Multimed Tools Appl 78, 35135–35156. https://doi.org/10.1007/s11042-019-08096-w

Bloomberg, M.R., Burden, A.M., 2006. Pedestrian Level of Service Study, Phase 1. NYC DCP Transportation Division, 167.

Capocchi, A., Vallone, C., Pierotti, M., Amaduzzi, A., 2019. Overtourism: A Literature Review to Assess Implications and Future Perspectives. Sustainability 11, 3303. https://doi.org/10.3390/su11123303

Dee, H.M., Velastin, S.A., 2008. How close are we to solving the problem of automated visual surveillance? Machine Vision and Applications 19, 329–343. https://doi.org/10.1007/s00138-007-0077-z

Fruin J.J., 1971. Pedestrian planning and design. Metropolitan Association of Urban Designers and Environmental Planners.

Fukuzaki, Y., Mochizuki, M., Murao, K., Nishio, N., 2015. Statistical analysis of actual number of pedestrians for Wi-Fi packet-based pedestrian flow sensing, in: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, UbiComp/ISWC'15 Adjunct. Association for Computing Machinery, New York, NY, USA, pp. 1519–1526. https://doi.org/10.1145/2800835.2801623

Gawande, U., Hajari, K., Golhar, Y., 2020. Pedestrian Detection and Tracking in Video Surveillance System:

Issues, Comprehensive Review, and Challenges, Recent Trends in Computational Intelligence. IntechOpen. https://doi.org/10.5772/intechopen.90810
HCM, 2000. Highway Capacity Manual. Transportation Research Board, Washington, D. C.
Hwang, Lee, Kim, 2019. Real-Time Pedestrian Flow Analysis Using Networked Sensors for a Smart Subway System. Sustainability 11, 6560. https://doi.org/10.3390/su11236560
Kaiser, M.S., Lwin, K.T., Mahmud, M., Hajializadeh, D., Chaipimonplin, T., Sarhan, A., Hossain, M.A., 2018. Advances in Crowd Analysis for Urban Applications Through Urban Event Detection. IEEE Trans. Intell. Transport. Syst. 19, 3092–3112. https://doi.org/10.1109/TITS.2017.2771746
Kalogianni, E., Sileryte, R., Lam, M., Zhou, K., Van der Ham, M., Van der Spek, S.C., Verbree, E., 2015. Passive WiFi monitoring of the rhythm of the campus. Proceedings of The 18th AGILE International Conference on Geographic Information Science; Geographics Information Science as an Enabler of Smarter Cities and Communities, Lisboa (Portugal), June 9-14, 2015; Authors version.
Kambatla, K., Kollias, G., Kumar, V., Grama, A., 2014. Trends in big data analytics. Journal of Parallel and Distributed Computing, Special Issue on Perspectives on Parallel and Distributed Processing 74, 2561–2573. https://doi.org/10.1016/j.jpdc.2014.01.003
Keith Still, G., 2000. crowd dynamics (PhD Thesis). University of Warwick.
Ken Victor Leonard, H., 2020. Kyoto and the Peril of Overtourism: Interview with Mayoral Candidate Murayama Shōei. URL https://www.nippon.com/en/japan-topics/c07701/kyoto-and-the-peril-of-overtourism-interview-with-mayoral-candidate-murayama-shoei.html
Kyritsis, D., 2017. The identification of road modality and occupancy patterns by Wi-Fi monitoring sensors as a way to support the “Smart Cities” concept: Application at the city centre of Dordrecht.
Lamba, S., Nain, N., 2017. Crowd Monitoring and Classification: A Survey, in: Advances in Computer and Computational Sciences. pp. 21–31. https://doi.org/10.1007/978-981-10-3770-2_3
Lee, S., 2021. Over Tourism Classification for Sightseeing Areas in Japan. Proceedings of The 3rd International Conference on Research in Social Sciences and Humanities
Lew, A., McKercher, B., 2006. Modeling Tourist Movements: A Local Destination Analysis. Annals of Tourism Research 33, 403–423. https://doi.org/10.1016/j.annals.2005.12.002
Li, J., Xu, L., Tang, L., Wang, S., Li, L., 2018. Big data in tourism research: A literature review. Tourism Management 68, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009
Liebig, T., Kemloh Wagoum, A.U., 2012. Modelling Microscopic Pedestrian Mobility Using Bluetooth, in: ICAART 2012 - Proceedings of the 4th International Conference on Agents and Artificial Intelligence. p. 275.
McKinsey & Company, & World Travel & Tourism Council, 2017. Coping with success: Managing overcrowding in tourism destinations.
Oklevik, O., Gössling, S., Hall, C.M., Steen Jacobsen, J.K., Grøtte, I.P., McCabe, S., 2019. Overtourism, optimisation, and destination performance indicators: a case study of activities in Fjord Norway. Journal of Sustainable Tourism 27, 1804–1824. https://doi.org/10.1080/09669582.2018.1533020
Orellana, D., Bregt, A.K., Ligtenberg, A., Wachowicz, M., 2012. Exploring visitor movement patterns in natural recreational areas. Tourism Management 33, 672–682. https://doi.org/10.1016/j.tourman.2011.07.010
Park, S., Xu, Y., Jiang, L., Chen, Z., Huang, S., 2020. Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data. Annals of Tourism Research 84, 102973. https://doi.org/10.1016/j.annals.2020.102973
Pechlaner, H., Innerhofer, E., Erschbamer, G., 2019. Overtourism: Tourism Management and Solutions. Routledge.
Peeters, P., Gössling, S., Klijs, J., Milano, C., Novelli, M., Dijkmans, C., Eijgelaar, E., Hartman, S., Heslinga, J., Isaac, R., Mitas, O., Moretti, S., Nawijn, J., Papp, B., Postma, A., 2018. Research for TRAN Committee - Overtourism: impact and possible policy responses 260.
Satake, K., Uno, N., Nakamura, T., 2019. Wi-Fi パケットセンサを用いた観測間隔を考慮した代表的観光周遊パターン分析 第 60 回土木計画学研究発表会・講演集 28-4.
Schauer, L., Werner, M., Marcus, P., 2014. Estimating crowd densities and pedestrian flows using wi-fi and bluetooth, in: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MOBIQUITOUS ’14. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, pp. 171–177. https://doi.org/10.4108/icst.mobiquitous.2014.257870
Timmermans, H. (Ed.), 2009. Pedestrian Behavior: Models, Data Collection and Applications. Emerald Group Publishing Limited. https://doi.org/10.1108/9781848557512

Traunmueller, M.W., Johnson, N., Malik, A., Kontokosta, C.E., 2018. Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities. Computers, Environment and Urban Systems 72, 4–12. https://doi.org/10.1016/j.compenvurbsys.2018.07.006

Turner, L., Ash, J., 1975. The Golden Hordes: International Tourism and the Pleasure Periphery. Constable Limited, London, UK.

UNWTO, 2018. Overtourism? Understanding and Managing Urban Tourism Growth beyond Perceptions. Madrid, Spain.

UNWTO (Ed.), 2004. Tourism Congestion Management at Natural and Cultural Sites. World Tourism Organization (UNWTO). https://doi.org/10.18111/9789284407637

Ward, C., Berno, T., 2011. Beyond social exchange theory: Attitudes Toward Tourists. Annals of Tourism Research 38, 1556–1569. https://doi.org/10.1016/j.annals.2011.02.005

Xia, C., Ciesielski, V., Arrowsmith, C., 2005. Data Mining of Tourists ' Spatio-temporal Movement Patterns- A Case Study on Phillip Island Jianhong ([WWW Document]. URL https://www.semanticscholar.org/paper/Data-Mining-of-Tourists-%E2%80%99-Spatio-temporal-Movement-Xia-Ciesielski/cbb5509a634e0264ee0cf6f71404a327f47c15d5 (accessed 4.5.22).

Zhan, B., Monekosso, D.N., Remagnino, P., Velastin, S.A., Xu, L.-Q., 2008. Crowd analysis: a survey. Machine Vision and Applications 19, 345–357. https://doi.org/10.1007/s00138-008-0132-4

Chapter 2: Related work and sensor specifications

This chapter aims to explore possible pedestrian data collection methods and explain why the Wi-Fi packet sensor was chosen as the data collection method in this research. The Wi-Fi packet sensor specifications is also explained in this chapter.

2.1 Pedestrian data acquisition methods

Pedestrian counting data is of great importance in many areas. For example, in a smart home, we can control the lighting, heating and cooling based on the number of people in the room. In a shopping mall, if we can know the number of people and their stay time in an area, we can analyse their consumption habits and preference, this will be helpful for selecting shop location. This information is also help for the design and management of public places such as tourist spot, bus stops, railway stations. Human movement behaviour research is also of increasing interest, particularly in the field of transportation planning. Movement data, indicating spatio-temporal characteristics of the flow of people, contain important basic information for understanding the behaviour of people and the formulation and evaluation of traffic countermeasures. In order to analyse and understand pedestrian behaviour, collecting pedestrian data is of great importance. A variety of data collection methods have been used to investigate pedestrian behaviour, such as field observations, controlled experiments and survey methods. The field observations aim to study the pedestrian behaviour in realistic, natural environments. Contrary to field observations, controlled experiments need the participants' movements in a controlled condition and a temporary experimental setup designed by the researchers.

Collecting pedestrian data from a field observation is not an easy work especially when the study area is huge and the target is great quantity. Many methodologies have been studied to collect pedestrian data. Daamen et al. (2016) classified the data collection methods for pedestrians, as shown in Table 2.1.

Table 2.1 Classification of data collection methods for pedestrians (Daamen et al., 2016)

		Measurement objective	
		Local	Global
Measurement perspective	Microscopic	Video Time-lapse Infrared Laser	Stalking Questionnaires GPS Bluetooth, Wi-Fi Mobile phone data
	Macroscopic	Manual counts Video Time-lapse Infrared Laser	Aerial observations GPS Bluetooth, Wi-Fi Mobile phone data

2.1.1 Conventional pedestrian data collection

Traditional pedestrian mobility monitoring methods include manual count, questionnaire survey, and cameras with image processing techniques (Li et al., 2014; Peters et al., 2010; Lam et al., 1995). As the name suggests, manual counting is a simple method that a person counts the number of individuals crossing designated sections. The accuracy of manual counting data can be relatively high, but it can only provide data only for limited time and locations since manual counting is labour intensive, and there is also a need for well-trained counting staff. Therefore, the manual counting appears powerless in the face of large crowds. One issue with surveys is that asking people to complete questionnaires can trigger survey fatigue and less faithful answers (Lee et al., 2016), and it often entails a great deal of preparatory work and post-data processing, high costs, and small sample sizes, which result in infrequent updates.

With the development of video recording and digital video recorder technology, researchers can collect pedestrian data through video cameras instead of going to the field. In the beginning, the video was taken firstly while counted manually. Later, thanks to the image processing technology, it can automatically count the number of pedestrians or even derive the crowd speed and density information (Wang et al., 2012; Favaretto et al., 2016). Lam et al. (1995) studied the pedestrian dynamics on six pedestrian facilities in Hong Kong. The walking speed and pedestrian count data were collected both by video cameras and on-site manual counts. They checked the speed-flow-density relationships of indoor and outdoor walkways, signalized crosswalks, light rail transit crosswalks, and stairways and built the basis for the development of pedestrian design standards for Hong Kong. However, the camera-based pedestrian detection method also has disadvantages as it requires good lighting conditions and the viewing area is limited. Besides, oldest methods for counting the amount of pedestrian flow have been using the top-view cameras, and most of the records are the tops of peoples' heads, therefore it is difficult to reconstruct individual movements across multiple camera views especially when there are a large number of pedestrians.

Although these methods can monitor the volume of people, they have several disadvantages, such as time consuming, limited coverage, high cost, and lack of movement information. In addition, it is difficult to obtain long-term data continually, and sometimes impractical, especially when the density of pedestrians is high.

2.1.2 Emerging pedestrian data source

By contrast, technological advancements have created a variety of data acquisition methods that require less effort and fewer resources yet produce a larger volume of data than traditional counting and surveying, as is the case with new technology-based data available through mobile devices. Emerging data sources have been explored to determine their usefulness with respect to measuring trip-making behaviour (Musa and Eriksson, 2012; Shoval and Ahas, 2016). When traffic volume over a long time period is needed, the automatic counting methods using diverse sensors such as pneumatic tubes, inductive loop detectors, infrared sensors, and radio beams can substitute for human data collectors (Lee and Sener, 2020). However, the installation and maintenance costs of these sensors are relatively high.

2.1.2.1 Different technologies for collecting pedestrian data

With the emergence of information and communication technology, smart technology has been able to mitigate many data collection issues and thus provide innovations to survey studies (Peters et al., 2010; Fukuda et al., 2017). These new technologies can capture high-resolution data on the behaviour of individual travellers in a large-scale population (Hasnat and Hasan, 2018). With the rapid expansion of smart mobile device ownership, new data sources have become available for studies on understanding crowd behaviour. Harari et al. (2016) provided an overview of the most common types of smartphone data and their application. The types of data include Bluetooth, GPS, Wi-Fi, and other log data. Musa and Eriksson (2012) used MAC address data for human movement-tracking technology. They described a system using Wi-Fi detection to passively track smartphone clients and presented a trajectory estimation method.

Shoval and Isaacson (2007) compared tourist tracking techniques that are based on GPS, land-based tracking systems, and hybrid solutions. They found that the time difference of arrival technique has a distinct advantage over GPS because it uses a light and easy-to-carry device, although GPS devices can obtain more accurate data. Martani et al. (2017) evaluated the reliability of two pedestrian monitoring systems; one utilises a downward-facing infrared depth sensor, and the other is based on a type of visible light (RGB) camera. The accuracy and limitations of both approaches under different conditions are discussed. it suggests that, although video surveillance has a good capture rate, the method is vulnerable to variation in lighting conditions, viewing angles, and weather conditions. The cost of video-based surveys is also relatively high. On the other hand, there are objections to the use of mobile network Global System for Mobile Communications log files due to privacy concerns (Utsch and Liebig, 2012).

Digital footprints are widely used in social mobility studies. One type of digital footprint is the photos uploaded on websites such as Flickr (an image hosting and video hosting service). Önder et al. (2016) analysed tagged photos on Flickr and demonstrated that these digital footprints could be used as a useful indicator of tourist numbers at a destination. Bermingham and Lee (2014) noted that numerous studies have used social media platforms to explore travellers' behaviour and trajectory patterns. By combining a collection of Flickr photos, they formulated the movement trails of an individual, called the trajectory. Spatio-temporal mining of the Queensland Flickr dataset uncovered interesting seasonal patterns along the east coast and local yearly patterns in Brisbane. Hasnat and Hasan (2018) presented a framework on how to use location-based data from social media (Twitter) to gather and analyse the travel behaviour of tourists. Some other researchers have also carried out studies using cellular data. Padrón and Hernández (2020) summarised how tracking techniques could be used in tourism research and to improve over-tourism management. To provide a structured and easy to follow evaluation of data characteristics, Table 2.2 presents an overview of different pedestrian data sources.

To summarise, most traditional pedestrian monitoring methods can just acquire small sample data because of the limitations of labour and cost, and the emerging data sources, especially smartphone-based method, have shown great potential for monitoring pedestrians. Studies using emerging data sources have focused on the volume estimation or time-based analysis, but the analysis of human movement based on space on an aggregation level is still not enough.

Table 2.2 Classification of pedestrian data sources

Pedestrian data sources	
Traditional	Emerging
Manual counting (human data collectors in the field or using video recordings)	Cell tower mobile phone positioning (e.g., AirSage)
Automated counting (e.g., pneumatic tubes, inductive loop detectors, infrared and radio beams)	Global positioning systems (e.g., INRIX)
Travel surveys (e.g., National Household Travel Survey)	Multi-app location-based service (e.g., Foursquare, Yelp, TripAdvisor, Facebook)
Interview surveys	App-based tracking (fitness/activity tracking apps)
Web-based surveys	Wi-Fi/Bluetooth

2.1.2.2 Wi-Fi and Bluetooth data

The development of information and communication technology has changed society and our life fundamentally, and we are living in a mobile information era now. As the smart mobile devices (smartphone, laptop) spread day by day, new streams of data are being generated and can be integrated and analysed to better understand pedestrian mobility patterns. A Wi-Fi packet sensor can record when and where the smart device carrier has been, which can be treated as a kind of sequential data for the analysis of human movement behaviour.

In 2020, 78.5% of the population in Japan used a smartphone, which constituted a significant increase from less than 64% in 2017 (Statista Research Department, 2021). The number of smartphone users worldwide today surpasses six billion and is forecasted to grow further by several hundred million in the next few years (S. O'Dea, 2022). Besides, wireless internet access has become a standard feature of smartphones, and each smartphone has a media access control (MAC) address that is unique to each device. The unique identifiers can be matched over space and time; therefore, the data are ideal for tracking devices. This information means that Wi-Fi probe request data sources are becoming more and more massive, and the Wi-Fi probe detection-based data collection method is becoming increasingly useful. Where point-to-point sensors are permanently installed, the amount of data collected can quickly become very large. The extracted information from such tracking data might be valuable and helpful for different kinds of use cases, such as crowd control, emergency situations, or just commercial purposes.

The use of media access control (MAC) data to track people has recently focused on applications to mass events, shopping centres, airports, train stations, and so on. Ferro and Potorti (2005) examined the difference between Bluetooth and Wi-Fi wireless protocols and concluded that Wi-Fi has a more extensive operating range, of up to 100 meters. Abedi et al. (2015) also investigated the effects of various antenna characteristics on pedestrian and cyclist travel time estimation using Bluetooth and Wi-Fi sensors. According to their results, the data collection rate from Wi-Fi-based sensors is almost 10 times theoretically and 8 times empirically larger than that of Bluetooth-based sensors. They also compared Wi-Fi and Bluetooth in terms of architecture, discovery time, popularity of use and signal strength. Similarly, Boehm et al. (2016) discussed whether Bluetooth and Wi-Fi sensors are suitable for reliable estimation of pedestrian volumes in urban areas. To test these sensors, several field tests were carried out under a non-motorised traffic condition (on two bridges). The received signal strength indicator (RSSI) was used as the main tool to select moving devices. They reported that data from Wi-Fi-based sensors perform better than Bluetooth-based sensors. Moreover, MAC address discovery time is important for efficiently collecting data in a short period of time. Lesani and Miranda-Moreno (2019) also developed a Bluetooth-Wi-Fi system to collect pedestrian data. They evaluated the performance of a system with three sensors at a pedestrian-only street at the McGill University campus over the course of 6 days. The ground truth data were obtained by manual counting from video recordings. They showed that the detection rate for Wi-Fi systems is 26%, while it is only 2.02% in the case of Bluetooth systems. Kurkcu and Ozbay (2017)examined pedestrian flows, wait times, and counts based on data collected from Wi-Fi and Bluetooth sensors. The developed methods were applied to data collected at a public transportation terminal using six sensors over the course of 2 months. They presented procedures to remove low-quality detections and improve the detection and counting performance of the devices. The theory of Wi-Fi and Bluetooth based data collection method are similar. Both are through detect the smart devices carried by users to track users. But Wi-Fi sensing technology has a more wide range and can detect more data.

Andión et al. (2018) studied a dataset collected during 1 year from nine Wi-Fi tracking sensors deployed on a university campus. Their data analyses included time and occupancy, people's positions, movements and identification of common behaviours, and a comparison between the actual data and the results collected from a video system at the main entrance of the university library. They reported that Wi-Fi tracking is more accurate than video camera systems and is cost-efficient. Their study illustrates how a low-cost Wi-Fi tracking system can be used under real-life conditions to improve the operation of monitored premises. Fukuzaki et al. (2014) developed a system that analysed pedestrian flow using Wi-Fi packet sensors. They confirmed that the approximate features of pedestrian flow could be analysed using their system and simple analytical methods with experiments in the laboratory and during an event at the

Osaka Electro-Communication University campus. Schauer et al. (2014) investigated the quality and feasibility of pedestrian flow estimations based on Wi-Fi and Bluetooth data captured in a realistic scenario at a German airport. Kalogianni et al. (2015) examined the rhythm of a university campus using 20 Wi-Fi monitors to collect data over 1 week at the Delft University of Technology. They focused on the occupancy, duration of stay and movement pattern at and between different facilities. Crawford et al. (2018) analysed the repeated trip behaviour of travellers in Wigan, a town in England, based on 1 year of data from 23 fixed Bluetooth sensors. They proposed a method for obtaining road user classifications based on their spatial and temporal variabilities. Abedi et al. (2014) presented the use of the MAC address data collection approach for the analysis of spatio-temporal dynamics of people in terms of shared space utilisation. Analysis of MAC address data in the university staff lounge provided clear statistics, such as utilisation frequency by staff, utilisation peak periods and time spent by staff. Wepulanon et al. (2019) proposed a method for bus passenger waiting time estimation using passive Wi-Fi data. They proposed a methodology to handle massive noise in Wi-Fi data and identified potential Wi-Fi records that could be derived from passenger's devices.

Ribeiro et al. (2020) developed a passive Wi-Fi tracking system and installed their sensory infrastructure in 19 buses to collect data related to public transport usage in the whole city. They analysed their data on a per-vehicle and per-stop basis and compared these against ground truth data (ticketing). Their study shows how collected data can be put to good use to improve the daily mobility experience involving sustainable mobility. Similar studies can be found in (Petre et al. 2017; Alekseev and William, 2019; Huang et al. 2019). Duives (2020) presented a large, though not comprehensive, overview of the studies tracking pedestrian with Wi-Fi or Bluetooth sensors, part of that is shown by Table 2.3.

Although Wi-Fi packet sensor data has been applied to analyse pedestrian behaviour, it is still not well developed in the current situation, there are still challenges facing it and the potential of using this emerging data source still needs to be broadened in various perspectives. There are still gaps when using Wi-Fi packet sensor technology, such as, how the environmental and installation conditions influence the detection capability of sensors, and the correlation between the sensor observation and actual pedestrian count is not clear. Moreover, how Wi-Fi sensing technology can be used to alleviate pedestrian crowding problems and study tourist behaviour is still limited.

Table 2.3 Summary of literature related to derivation traffic state using Wi-Fi/Bluetooth sensors

Reference	Sensor type	Setting
(O'Neill et al., 2006)	Bluetooth	City streets of Bath
(Miyaki et al., 2007)	Wi-Fi	City streets and Uni. of Tokyo
(Musa and Eriksson, 2012)	Wi-Fi	City streets of Chicago
(Kostakos et al., 2013)	Wi-Fi	City streets of Oulu, Finland
(Danalet et al., 2014)	Wi-Fi	Campus
(Schauer et al., 2014b)	Bluetooth and Wi-Fi	Security gates airport
(Fukuzaki et al., 2015b)	Wi-Fi	Shopping mall
(Ma et al., 2015)	Wi-Fi	University building
(Daamen et al., 2016)	Wi-Fi	Nautical event
(Hoogendoorn et al., 2016)	Wi-Fi	Nautical event
(Bellini et al., 2017)	Wi-Fi	City streets San Francisco
(Fang and Hong, 2017)	Wi-Fi	University campus Dartmouth
(Potortì et al., 2018)	Wi-Fi	Building
(Duives et al., 2018)	Wi-Fi	Music event

2.1.3 Pros and cons of Wi-Fi packet sensor data

Judging from these studies, Wi-Fi-based methods have clear advantages. As smartphones equipped with Wi-Fi modules are ubiquitous, the cost of deploying a Wi-Fi-based crowd-tracking system is rather small, and such schemes enable us to obtain long-term and continuous counts. Moreover, there is no need for direct interaction with data donors, people carrying Wi-Fi devices do not need to install any apps for such data to be collected, and the data output by the system is easy to process. The other advantage of this method is the installation location of the Wi-Fi packet sensors is very flexible. However, like authors in Al Ameen (2012) say that any wireless system has some inherent technical vulnerabilities and limitations. Sending data out from the Wi-Fi packet sensor through wireless media can pose threats to the privacy of an individual. The concerns for privacy have been investigated by Nishida et al. (2018). According to their study, the device-specific information contained in the packets cannot be used to identify an individual by itself, but it can be maliciously linked to personal information, such as by tracking a target individual and obtaining the MAC address and thus behavioural tracking of individuals may be performed. Therefore, they convert the obtained MAC address to anonymised MAC address using a one-way hash function in the sensor, and change the salt of hash function weekly.

2.2 Overview of Wi-Fi packet sensor based data collection system

Wireless Fidelity (Wi-Fi) also known IEEE 802.11 (Crow et al., 1997; Willig, 2003; Ferro and Potorti, 2005; Gast, 2005) is designed for wireless local area network connections. The MAC address is a unique identifier assigned to hardware for communication on the internet, it is like the address in the internet. Electronic devices, such as smartphones, tablets, and computers with Wi-Fi enabled periodically transmit so-called 'probe requests,' even when the device is not associated with a network. Probe requests include a MAC address that is unique for each device, and thus the Wi-Fi packet sensor can be used to identify the movement of its holders. To protect the user's privacy, some smart device manufacturers have developed technology to randomise MAC addresses. These include the producers of devices running Android or iOS operating systems. However, it can distinguish between real and randomised MAC addresses using certain technical means (Martin et al., 2017). It is therefore possible to count and track pedestrians and by detecting the devices they are carrying. Figure 2.1 shows an overview of Wi-Fi packet-based tracking systems. An anonymised MAC address probe (Wi-Fi packet) sensor can be used to detect probe requests from smart devices and upload them to cloud servers after anonymisation. To protect device owners' privacy, the sensor anonymises MAC addresses to A-MAC addresses.

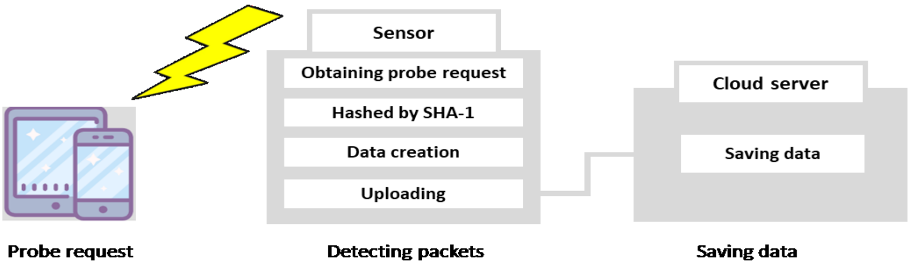

Figure 2.1 Overview of Wi-Fi packet-based tracking system

The Wi-Fi packet sensors used in this study were developed by the Japan Research Institute for Social Systems, and the data recorded by the sensors is shown in Table 2.4. Figure 2.2 represents a screen shot from data base.

Table 2.4 Data information acquired from Wi-Fi packet sensor

ID	Record number
UNIXTIME	Internal time (it is possible to acquire decimal points in seconds)
TIMESTAMP	Date and time when the packet was captured (converted to seconds)
AMPID	Sensor ID Which sensor captured the packet,
AMAC	MAC address after anonymisation --a unique device identifier
OUI	MAC address vendor code (Used for judgment of randomized MAC address etc.)
RSSI	Received signal strength indicator

1	ID	UNIXTIME	TIMESTAMP	AMPID	AMAC	SC	OUI	RSSI
2	49777203	1510412405	12/11/2017 00:00	AMPM17-GN009	a69008e403028a8e40a789be3568f9e05cb15b84	3204	FC62B9	-86
3	49777271	1510412401	12/11/2017 00:00	AMPM17-GN001	eea72a9a7a589816fba8b2ead2cc0f0c83102f8d	597	3A75D5	-86
4	49777272	1510412403	12/11/2017 00:00	AMPM17-GN001	96095c751045c83e49311f32143111b55535442e	127	4.04E+38	-86
5	49777273	1510412404	12/11/2017 00:00	AMPM17-GN001	67f520b5aa1db39d40a70581cf9d71954db045b5	1862	22C40B	-90

Figure 2.2 Screen shot from the data base

References

Abedi, N., Bhaskar, A., Chung, E., 2014. Tracking spatio temporal movement of human in terms of space utilization using media access control address data. Applied Geography 51, 72–81.

Abedi, N., Bhaskar, A., Chung, E., Miska, M., 2015. Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. Transportation Research Part C: Emerging Technologies 60, 124–141. https://doi.org/10.1016/j.trc.2015.08.010

Alekseev, N., Lam, W.H.K., 2019. Estimation of Pedestrian Flow Based on Wi-Fi Data and Video Cameras. Journal of the Eastern Asia Society for Transportation Studies 13, 93–101. https://doi.org/10.11175/easts.13.93

Andión, J., Navarro, J.M., López, G., Álvarez-Campana, M., Dueñas, J.C., 2018. Smart Behavioral Analytics over a Low-Cost IoT Wi-Fi Tracking Real Deployment. Wireless Communications and Mobile Computing 2018, e3136471. https://doi.org/10.1155/2018/3136471

Bellini, P., Cenni, D., Nesi, P., Paoli, I., 2017. Wi-Fi based city users' behaviour analysis for smart city. Journal of Visual Languages & Computing 42, 31–45. https://doi.org/10.1016/j.jvlc.2017.08.005

Bermingham, L., Lee, I., 2014. Spatio-temporal Sequential Pattern Mining for Tourism Sciences. Procedia Computer Science, 2014 International Conference on Computational Science 29, 379–389. https://doi.org/10.1016/j.procs.2014.05.034

Boehm, M., Ryeng, E., Haugen, T., 2016. Evaluating the Usage of Wi-Fi and Bluetooth Based Sensors for Pedestrian Counting in Urban Areas, in: Undefined. Presented at the European Transport Conference 2016Association for European Transport (AET).

Crawford, F., Watling, D.P., Connors, R.D., 2018. Identifying road user classes based on repeated trip behaviour using Bluetooth data. Transportation Research Part A: Policy and Practice 113, 55–74. https://doi.org/10.1016/j.tra.2018.03.027

Crow, B.P., Widjaja, I., Kim, J.G., Sakai, P.T., 1997. IEEE 802.11 Wireless Local Area Networks. IEEE Communications Magazine 35, 116–126. https://doi.org/10.1109/35.620533

Daamen, W., Yuan, Y., Duives, D.C., Hoogendoorn, S.P., 2016. Comparing three types of real-time data collection techniques: Counting cameras, Wi-Fi sensors and GPS trackers. Proceedings of Pedestrian and Evacuation Dynamics 2016.

Danalet, A., Farooq, B., Bierlaire, M., 2014. A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. Transportation Research Part C: Emerging Technologies 44, 146–170. https://doi.org/10.1016/j.trc.2014.03.015

Duives, D.C., Tim van Oijen, Serge P, H., 2020. Enhancing Crowd Monitoring System Functionality through Data Fusion: Estimating Flow Rate from Wi-Fi Traces and Automated Counting System Data. Sensors 20, 1–25.

Duives, D.C., van Oijen T, Daamen, W., Hoogendoorn, S., 2018. Data-driven state estimation using GPS traces: Density estimation @ Mysteryland 2017, in: Proceedings of the 9th Conference on Pedestrian and Evacuation Dynamics. Lund, Sweden, pp. 21–24.

Fang, T., Hong, X., 2017. Discovering Meaningful Mobility Behaviors of Campus Life from User-Centric WiFi Traces, in: Proceedings of the SouthEast Conference, ACM SE '17. Association for Computing Machinery, New York, NY, USA, pp. 76–80. https://doi.org/10.1145/3077286.3077306

Favaretto, R.M., Dihl, L.L., Musse, S.R., 2016. Detecting crowd features in video sequences, in: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Presented at the 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 201–208.

https://doi.org/10.1109/SIBGRAPI.2016.036
Ferro, E., Potorti, F., 2005. Bluetooth and Wi-Fi wireless protocols: a survey and a comparison. IEEE Wireless Communications 12, 12–26. https://doi.org/10.1109/MWC.2005.1404569
Fukuda, D., Kobayashi Hana, Nakanishi Wataru, Suga Yoshiki, Kerkritt, S., Kasem, C., 2017. Estimation of Paratransit Passenger Boarding/Alighting Locations Using Wi-Fi based Monitoring: Results of Field Testing in Krabi City, Thailand. Journal of the Eastern Asia Society for Transportation Studies 12, 2151–2169. https://doi.org/10.11175/easts.12.2151
Fukuzaki, Y., Mochizuki, M., Murao, K., Nishio, N., 2015. Statistical analysis of actual number of pedestrians for Wi-Fi packet-based pedestrian flow sensing, in: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, UbiComp/ISWC'15 Adjunct. Association for Computing Machinery, New York, NY, USA, pp. 1519–1526. https://doi.org/10.1145/2800835.2801623
Fukuzaki, Y., Mochizuki, M., Murao, K., Nishio, N., 2014. A pedestrian flow analysis system using Wi-Fi packet sensors to a real environment, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp '14 Adjunct. Association for Computing Machinery, New York, NY, USA, pp. 721–730. https://doi.org/10.1145/2638728.2641312
Gast, M.S., 2005. 802.11 Wireless Networks: The Definitive Guide, 2nd Edition. O'Reilly Media, Inc.
Harari, G.M., Lane, N.D., Wang, R., Crosier, B.S., Campbell, A.T., Gosling, S.D., 2016. Using smartphones to sollect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect Psychol Sci 11, 838–854. https://doi.org/10.1177/1745691616650285
Hasnat, M.M., Hasan, S., 2018. Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. Transportation Research Part C: Emerging Technologies 96, 38–54. https://doi.org/10.1016/j.trc.2018.09.006
Hoogendoorn, S.P., Daamen, W., Duives, D.C., Yuan, Y., 2016. Estimating travel times using Wi-Fi sensor data. TRISTAN 2016.
Huang, B., Mao, G., Qin, Y., Wei, Y., 2019. Pedestrian Flow Estimation Through Passive WiFi Sensing. IEEE Transactions on Mobile Computing 20, 1529–1542. https://doi.org/10.1109/TMC.2019.2959610
Kalogianni, E., Sileryte, R., Lam, M., Zhou, K., Van der Ham, M., Van der Spek, S.C., Verbree, E., 2015. Passive WiFi monitoring of the rhythm of the campus, in: The 18th AGILE International Conference on Geographic Information Science. Agile.
Kostakos, V., Ojala, T., Juntunen, T., 2013. Traffic in the Smart City: Exploring City-Wide Sensing for Traffic Control Center Augmentation. IEEE Internet Computing 17, 22–29. https://doi.org/10.1109/MIC.2013.83
Kurkcu, A., Ozbay, K., 2017. Estimating Pedestrian Densities, Wait Times, and Flows with Wi-Fi and Bluetooth Sensors. Transportation Research Record 2644, 72–82. https://doi.org/10.3141/2644-09
Lam, Morrall, J., Ho, H.H., 1995. Pedestrian flow characteristics in Hong Kong. Transportation Research Record 1487, 56–62.
Lee, K., Sener, I.N., 2020. Emerging data for pedestrian and bicycle monitoring: Sources and applications. Transportation Research Interdisciplinary Perspectives 4, 100095. https://doi.org/10.1016/j.trip.2020.100095
Lee, R.J., Sener, I.N., Mullins, J.A., 2016. An evaluation of emerging data collection technologies for travel demand modeling: from research to practice. Transportation Letters 8, 181–193. https://doi.org/10.1080/19427867.2015.1106787
Lesani, A., Miranda-Moreno, L., 2019. Development and Testing of a Real-Time WiFi-Bluetooth System for Pedestrian Network Monitoring, Classification, and Data Extrapolation. IEEE Transactions on Intelligent Transportation Systems 20, 1484–1496. https://doi.org/10.1109/TITS.2018.2854895
Li, T., Huan Chang, Meng Wang, Bingbing Ni, Richang Hong, Shuicheng Yan, 2014. Crowded Scene Analysis: A Survey. IEEE Trans. Cir. and Sys. for Video Technol. 25, 367–386. https://doi.org/10.1109/TCSVT.2014.2358029
Ma, W., Zhu, X., Huang, J., Shou, G., 2015. Detecting Pedestrians Behavior in Building Based on Wi-Fi Signals, in: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). pp. 1–8. https://doi.org/10.1109/SmartCity.2015.38
Martani, C., Stent, S., Acikgoz, S., Soga, K., Bain, D., Jin, Y., 2017. Pedestrian monitoring techniques for crowd-flow prediction. Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction 170, 17–27. https://doi.org/10.1680/jsmic.17.00001
Martin, J., Mayberry, T., Donahue, C., Foppe, L., Brown, L., Riggins, C., Rye, E.C., Brown, D., 2017. A Study of MAC Address Randomization in Mobile Devices and When it Fails. Proceedings on Privacy Enhancing Technologies.

Miyaki, T., Yamasaki, T., Aizawa, K., 2007. Tracking Persons using Particle Filter Fusing Visual and Wi-Fi Localizations for Widely Distributed Camera, in: 2007 IEEE International Conference on Image Processing. Presented at the 2007 IEEE International Conference on Image Processing, p. III-225-III–228. https://doi.org/10.1109/ICIP.2007.4379287

Musa, A.B.M., Eriksson, J., 2012. Tracking unmodified smartphones using wi-fi monitors, in: Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, SenSys ’12. Association for Computing Machinery, New York, NY, USA, pp. 281–294. https://doi.org/10.1145/2426656.2426685

Onder, I., Koerbitz, W., Hubmann-Haidvogel, A., 2016. Tracing Tourists by Their Digital Footprints: The Case of Austria. Journal of Travel Research 55. https://doi.org/10.1177/0047287514563985

O’Neill, E., Kostakos, V., Kindberg, T., Fatah gen. Schieck, A., Penn, A., Fraser, D., Jones, T., 2006. Instrumenting the City: Developing Methods for Observing and Understanding the Digital Cityscape. p. 332.

Padrón-Ávila, H., Hernández-Martín, R., 2020. Tourist Tracking Techniques as a Tool to Understand and Manage Tourism Flows, in: Séraphin, H., Gladkikh, T., Vo Thanh, T. (Eds.), Overtourism: Causes, Implications and Solutions. Springer International Publishing, Cham, pp. 89–105. https://doi.org/10.1007/978-3-030-42458-9_6

Peters, M., Piazolo, F., Köster, L., Promberger, K., 2010. The Deployment of Intelligent Local-Based Information Systems (ilbi): A Case Study of the European Football Championship 2008. Journal of Convention & Event Tourism 11, 18–41. https://doi.org/10.1080/15470141003587574

Petre, A.-C., Chilipirea, C., Baratchi, M., Dobre, C., van Steen, M., 2017. WiFi Tracking of Pedestrian Behavior, in: Xhafa, F., Leu, F.-Y., Hung, L.-L. (Eds.), Smart Sensors Networks, Intelligent Data-Centric Systems. Academic Press, pp. 309–337. https://doi.org/10.1016/B978-0-12-809859-2.00018-8

Potortì, F., Crivello, A., Girolami, M., Barsocchi, P., Traficante, E., 2018. Localising crowds through Wi-Fi probes. Ad Hoc Networks 75–76, 87–97. https://doi.org/10.1016/j.adhoc.2018.03.011

Ribeiro, M., Galvão, B., Prandi, C., Nunes, N., 2020. Passive Wi-Fi Monitoring in Public Transport: A case study in the Madeira Island, in: Proceedings of TRA2020, the 8th Transport Research Arena: Rethinking Transport Towards Clean and Inclusive Mobility.

S. O’Dea., 2022. Smartphone users worldwide 2016-2021 [WWW Document]. Statista. URL https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

Schauer, L., Werner, M., Marcus, P., 2014. Estimating crowd densities and pedestrian flows using wi-fi and bluetooth, in: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. pp. 171–177. https://doi.org/10.4108/ICST.MOBIQUITOUS.2014.257870

Shoval, N., Ahas, R., 2016. The use of tracking technologies in tourism research: the first decade. Tourism Geographies 18, 587–606. https://doi.org/10.1080/14616688.2016.1214977

Shoval, N., Isaacson, M., 2007. Tracking tourists in the digital age. Annals of Tourism Research 34, 141–159. https://doi.org/10.1016/j.annals.2006.07.007

Statista research department, 2021. Smartphone penetration rate in Japan 2017-2026 [WWW Document]. Statista. URL https://www.statista.com/statistics/275102/share-of-the-population-to-own-a-smartphone-japan/

Utsch, P., Liebig, T., 2012. Monitoring microscopic pedestrian mobility using bluetooth. Presented at the Eighth International Conference On Intelligent Environments, pp. 173–177. https://doi.org/10.1109/IE.2012.32

Wang, B., Ye, M., Li, X., Zhao, F., Ding, J., 2012. Abnormal crowd behavior detection using high-frequency and spatio-temporal features. Machine Vision and Applications 23, 501–511. https://doi.org/10.1007/s00138-011-0341-0

Wepulanon, P., Sumalee, A., Lam, W.H.K., 2019. Temporal Signatures of Passive Wi-Fi Data for Estimating Bus Passenger Waiting Time at a Single Bus Stop. IEEE Transactions on Intelligent Transportation Systems 21, 3366–3376. https://doi.org/10.1109/TITS.2019.2926577

Willig, A., 2003. An architecture for wireless extension of PROFIBUS, in: IECON’03. 29th Annual Conference of the IEEE Industrial Electronics Society (IEEE Cat. No.03CH37468). Presented at the IECON’03. 29th Annual Conference of the IEEE Industrial Electronics Society (IEEE Cat. No.03CH37468), pp. 2369-2375 Vol.3. https://doi.org/10.1109/IECON.2003.1280615

Chapter 3: Fundamental analysis on the Wi-Fi packet sensor based data collecting system

3.1 Introduction and research objective

Understanding pedestrian flow in large public buildings, such as airports, train stations, and shopping malls, is a significant challenge for the people running and managing such buildings. Systems that can identify customer densities can support the control and management of people flows and thus reduce travel time and management costs. Such systems can help determine people flow-control strategies, e.g., by closing or opening additional doors, ticketing booths and/or control gates. If the people flow and density can be observed automatically in real time, customers can be informed about the degree of congestion at their desired destination and certain pedestrian flows can be led through less-crowded areas to save time or for better comfort. Furthermore, such crowd information is also potentially useful for commercial purposes. Pedestrian and vehicle volumes are also the key criteria used to evaluate road network use. If the density and flow of pedestrians and vehicles can be observed, we can provide better services and cost savings, reduce air pollution, and so on. These types of data can also be used for urban design, transportation planning, tourist behaviour analyses, evacuation planning, and many other purposes. In the era of data-driven research, there is an interesting trend toward developing crowd-behaviour models using real-world data, and the usage of Wi-Fi packet sensors has recently attracted the attention of researchers. However, many researchers use sensor data directly, without considering differences between the detection capabilities of sensors, but these may vary with the environment. Hence, this chapter explores how the conditions surrounding Wi-Fi packet sensors influence observation results. I further attempt to estimate vehicle and pedestrian flow volume and explore whether it can estimate the pedestrian attribute type based on Wi-Fi packet sensor data.

The remainder of this chapter is organised as follows. In Section 3.2, a brief overview of current research in this topic is provided. For the purpose of exploring the factors affecting the observation results of the Wi-Fi packet sensor, section 3.3 describes the experimental setup in the laboratory. To check how accurately pedestrian and traffic flow can be quantified, section 3.4 describes the campus experiment. The clustering analysis presented in section 3.5 is to investigate whether it could identify different types of smart device users based on the Wi-Fi packet sensor data. Finally, I conclude the paper and summarise future work in Section 3.6.

3.2 **Related research**

There have been many studies that have focused on Wi-Fi-based crowd-tracking systems. Among these, Musa and Eriksson (2012)may have been the first to use MAC address data for human movement-tracking technology. They described a system using Wi-Fi detection to passively track smartphone clients and presented a trajectory estimation method. Fukuzaki et al. (2014)developed a system that analyses pedestrian flow using Wi-Fi packet sensors. According to their results, the Wi-Fi probe request frame transmission interval is between 30 and 120 s (depending on the device), and the RSSI is proportional to the distance between the sensor and the device, as long as that distance is less than 15 metres. They also carried out experiments in the lab and during an event at the Osaka Electro-Communication University campus. They confirmed that they can analyse the approximate features of pedestrian flow using their system and simple analytical methods. Fukuzaki et al. (2015) continued to study the extent to which the actual number of pedestrians can be estimated based on Wi-Fi detection data. They carried out a 2-month field experiment in a shopping mall and calculated a coefficient for estimating the actual number of people within a mall by comparing data obtained from Wi-Fi packet sensors to data collected from motion detectors. Based on these data, they reported a recognition rate of 29.3% on weekdays and 35.6% on holidays and weekends. Kalogianni et al. (2015)examined the rhythm of a university campus using 20 Wi-Fi monitors to collect data over the course of 1 week at the Delft University of Technology. They focused on the user’s occupation, duration of stay, and movement pattern at and between different facilities. Andión et al. (2018)studied a dataset collected over the course of 1 year from nine Wi-Fi tracking sensors deployed in a university campus. Their data analyses included time and occupancy, people’s positions, movements, and common behaviours, and a comparison between the actual data and the results collected from a video system at the main entrance of the university library. They reported that Wi-Fi tracking is more accurate than video camera systems, while also being cost-efficient. Their study illustrates how a low-cost Wi-Fi tracking system can be used under real-life conditions to improve the operation of monitored premises.

Wi-Fi probe requests are a type of electromagnetic wave and factors such as the environment, obstacles, distance between the device and the sensor, and antenna gain will affect the detection rates of Wi-Fi packet sensors. Other factors that affect wireless transmission include attenuation distortion, free space loss, noise, atmospheric absorption, multipath, and refraction (Xhafa et al., 2017). Some of the studies mentioned above investigated the influence of antenna gain on Wi-Fi packet sensor observation performance but did not consider the basic characteristics of the sensors. The researchers carried out their experiments under the assumption that all sensors have the same detection capacity. However, the observation performance of sensors varies with the request transmission characteristics of the Wi-Fi probe, even when the sensors have the same type of antenna. In this chapter, I attempt to fill this gap. I carried out experiments to evaluate the factors that influence detection performance, estimated vehicle and pedestrian flows, and compared our estimates to known ground truth data.

3.3 **Laboratory experiment**

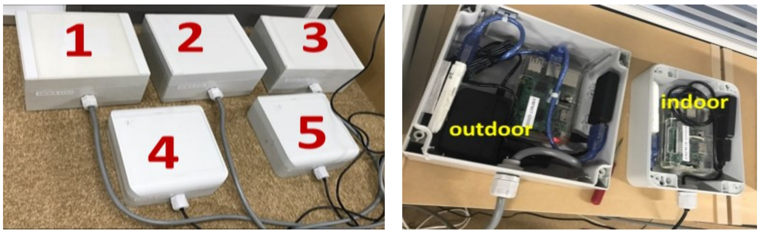

Figure 3.1 Wi-Fi packet sensors

In this chapter, experiments were carried out with the five sensors shown in Figure 3.1 (1,2,3 outdoor-type; 4,5 indoor-type). The outdoor-type sensors were waterproof and had bigger containers (160 × 160 × 90 mm, 840 g), whereas the indoor-type sensors had smaller containers (160 × 130 × 60 mm, 490 g) and electricity was provided via a USB socket. I first investigated the factors that impact the detection results, and then evaluated their vehicle and pedestrian detection abilities by carrying out experiments at the Gifu University Campus. First, the devices were installed at similar locations so that their detection tendency could be investigated. Then they were placed at different heights, as shown in Figure 3.2. The layout of the sensors was changed about every 6 days. I investigated the influence of height (high, middle, or low), sensor type (outdoor or indoor), and sensor ID (s1 to s5) on detection properties. Table 3.1 summarises the time and location data for the five sensors, with the number in the table indicating the sensor ID (s1 to s5).

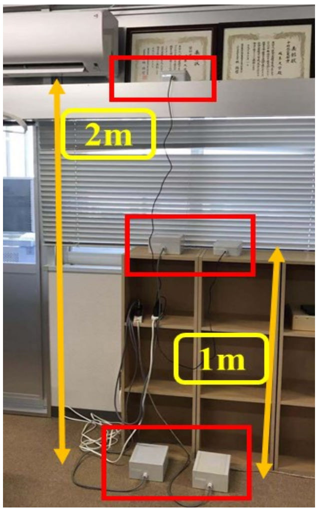

Figure 3.2 Layout of Wi-Fi packet sensors for height test

Table 3.1 Layout of the Wi-Fi packet sensors

Location / Time period	High (h3)	Middle (h2)	Low (h1)
6/29 17:00-7/5 15:00	s5, s4	s2	s3, s1
7/5 15:00-7/11 12:00	s3	s4, s5	s2, s1
7/11 12:00-7/18 11:00	s2, s1	s3	s4, s5
7/18 11:00-7/24 13:00	s4	s1, s5	s2, s3
7/24 15:00-7/31 23:00	s5, s3	s4, s1	s2

From Table 3.1, we can see that I collected height test data from 2018/6/29 17:00 to 2018/7/31 23:00. The analysis of variance (ANOVA) test was carried out to investigate the statistical differences between the conditions tested. The number of detected AMACs was aggregated per hour, and the variation in this number was calculated. Before making these calculations, I deleted the randomised AMAC addresses. There was a power cut during the height test in the period (7/24 13:50–7/24 14:20). Therefore, I deleted the data covering the 2-hour period during which this outage occurred (7/24 13:00–7/24 15:00).

Table 3.2 Analysis of variance (ANOVA) results

Factors	DF	F value	P value	CR
Sensor Type	1	262.561*	$<2\times 10^{-16}$	6.03%
Sensor ID	3	6.873*	0.00013	0.47%
Height	2	178.232*	$<2\times 10^{-16}$	8.18%
Sensor Type: Height	2	9.613*	6.85×10^{-5}	0.47%
Sensor ID: Height	4	70.795*	$<2\times 10^{-16}$	0.25%
Error	2			84.6%
Total	14			100%

*CR, contribution rate; DF, degree of freedom *: 0.1% significance*

The ANOVA test results are summarised in Table 3.2. The contribution rate (CR) represents the percentage contribution of each factor. The data suggest that the CR of the error accounted for 84.6% of the total variation. The major factor affecting the variation in the number of AMAC observations was the change in the number of detected devices located around the sensor, and this varied with respect to time and date, in accordance with the density of people in the area. These results suggest that about 15% of the variation can be explained by the sensor-type, sensor ID, and installation height, which means that the location of the sensor should be selected carefully to avoid differences in detection ability. Looking at the influence of individual factors, height was the most influential, explaining approximately 53% (=8.18%/15.4%) of the variation due to the sensor installation location. Sensor type was the second most influential factor, explaining 39% (=6.03%/15.4%) of the variation. Although sensor ID and the two interaction effects were statistically significant, they only explained a small amount of variation. Hence, these factors can be neglected for convenience in experimental design.

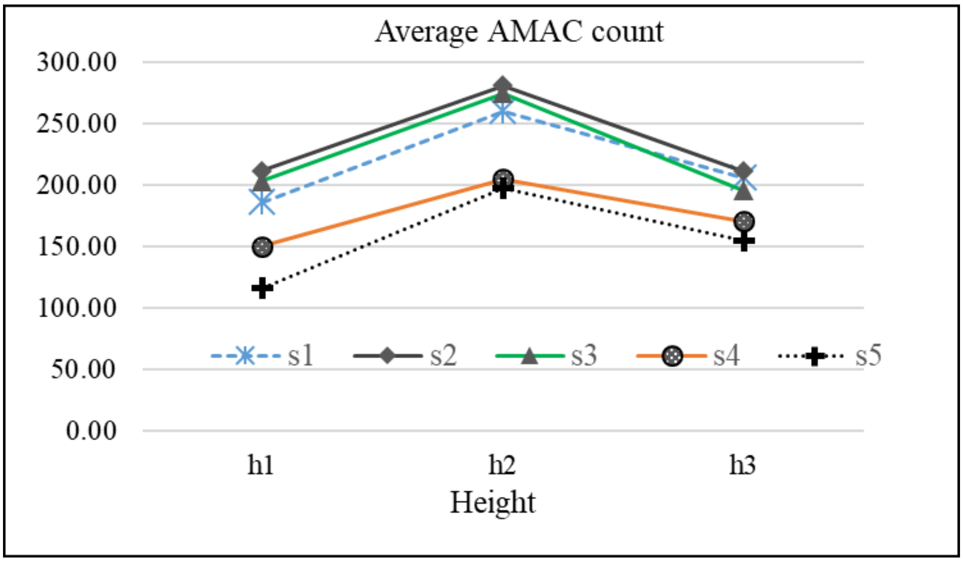

Figure 3.3 Average AMAC counts from sensors at different locations

To confirm this, Figure 3.3 shows the relationship between the average counts of the sensors at each location. We can see that the number of observations was largest when the sensors were at the middle height. In addition, the outdoor-type sensors (s1, s2, and s3) performed better than the indoor-type sensors (s4 and s5). In cases where s1 and s3 were installed at h1 and h2, s3 detected more AMACs than s1, but s1 detected more AMACs when they were installed at h3. Hence, the sensor height influences the detection capacity of Wi-Fi packet sensors. This can also be seen from Table 3.2. The relationship between sensor ID and height is also statistically significant. From Figures 3.4 and 3.5, we can see that the total AMAC counts and AMAC counts per day were smaller in the case of indoor-type sensors (s4 and s5) than outdoor-type sensors (s1, s2 and s3) most of the time.

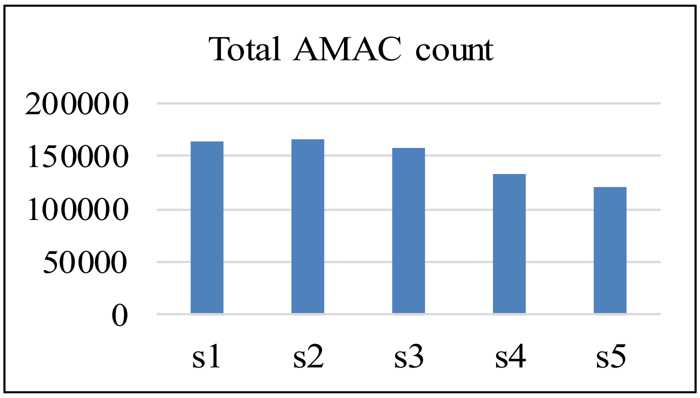

Figure 3.4 Total AMAC counts by sensor

Figure 3.5 shows the daily AMAC counts, from which we can see that the observations depend on whether the day is on a weekday or a weekend (Saturdays are coloured in blue, and Sundays and holidays are coloured in red). The vertical red lines show the height change timing. We can also conclude that there are differences between the detection abilities of indoor and outdoor sensors. Thus, the results of in-laboratory experiment suggest that the height and type of Wi-Fi packet sensor influence their detection capacity, so we should consider these factors when using Wi-Fi packet sensor data to estimate traffic flow.

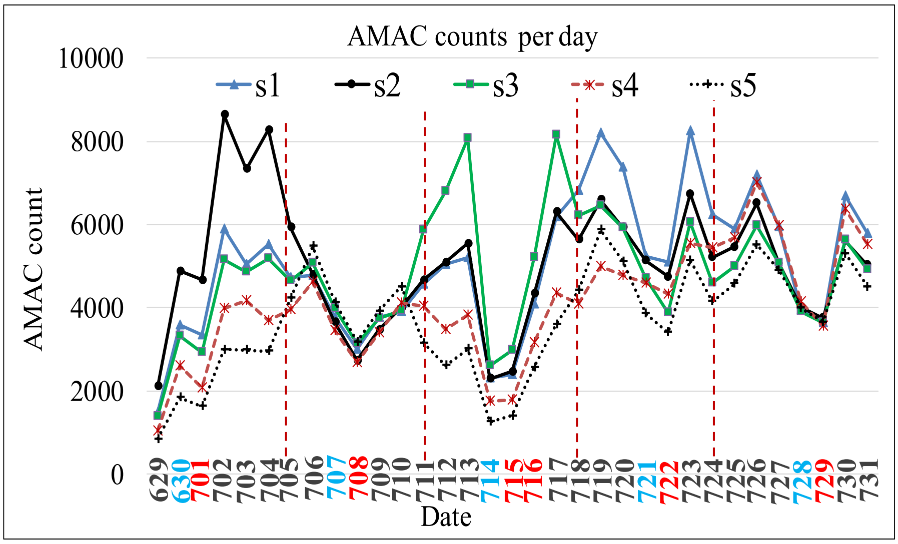

Figure 3.5 AMAC counts of Wi-Fi packet sensor data per day

The reason why the Wi-Fi packet sensors at the middle height had the best performance maybe some Wi-Fi probe requests were absorbed by the ceiling or floor or obstructed by the things around. The reason why the out-door type sensors performed better than the in-door type sensors may be that the outdoor-type sensor has a longer power cable with a transformer inside the container. The longer power cable just like an antenna may contribute to the detection capability of the out-door type sensor.

3.4 **Campus experiment**

A second experiment was carried out to verify the detection capabilities of sensors when they are used to quantify the number of vehicles and pedestrians. Fukuda et al. (2017) tried to estimate the boarding and alighting locations and the number of paratransit passengers using Wi-Fi and GPS data. They installed both Wi-Fi scanner and GPS logger in the paratransit vehicles to collect data. But I focus on the observation of all pedestrians using only fixed Wi-Fi packet sensors. Observation experiments were carried out at Gifu University Campus to quantify vehicles (2018/11/05–2018/12/10) and pedestrians (2019/1/11–2019/2/12).

The layout of the Wi-Fi packet sensors and their locations are shown in Figures. 3.6 and 3.7, respectively. V1–V5 and P1–P5 represent the sensors observing vehicles and pedestrians, respectively. I installed the sensors at locations that will be passed by many vehicles and pedestrians.

As shown in Figure 3.6, I installed V2, V3, and V5 at the university gates because vehicles must enter and exit the campus through these gates. V1 and V4 are near parking places, so vehicles often pass these locations. Due to power supply limitations, I had to place sensors V1 and V4 inside a building, and the distances from the road to V1 and V4 are 60 and 35 metres, respectively. To observe pedestrians, I placed the sensors where vehicles are prohibited so that only pedestrians and cyclists would be observed. At the same time, I carried out a video survey to record the vehicles as they passed by, and a manual counting survey to record the number of pedestrians. There are monitoring cameras at the entry gates of the university (V2, V3, and V5). I obtained these videos from the University, and also recorded by our own videos at V1 and V4 (2018/12/05 8:00–18:00). To observe pedestrians, I set an imaginary cordon line for each sensor and manually counted the pedestrians who crossed it (on 2019/2/6 for P1 and P4, and on 2019/2/7 for P2, P3 and P3). I used the number of vehicles from the video survey and number of pedestrians from the manual counting survey as ground truth data. Comparing the Wi-Fi packet sensor data to the ground truth data enabled us to estimate the detection rate of each sensor.

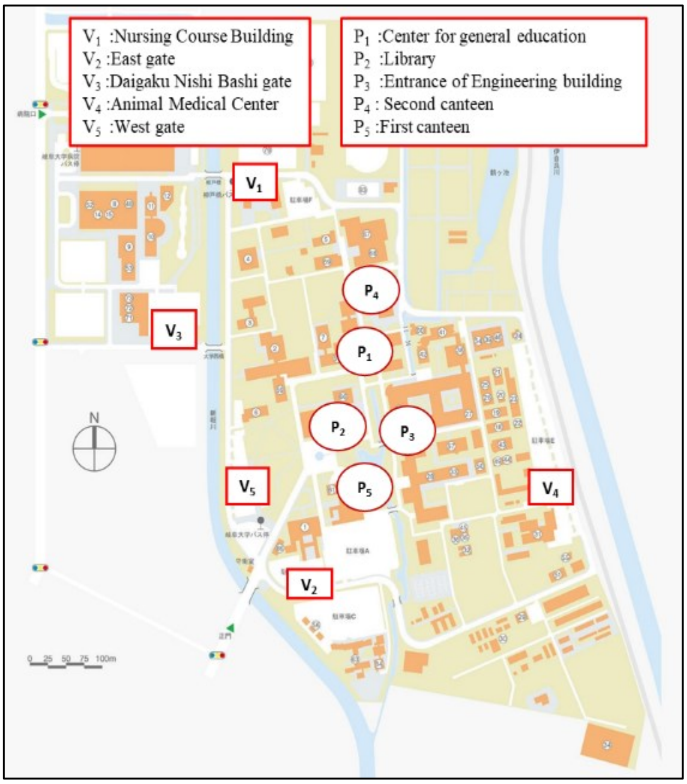

Figure 3.6 Layout of Wi-Fi packet sensors

V1: Refreshment corner

V2: Guard's room table

V3: On the ground

V4: On top of a shelf

V5: Guard's room table

P1: Passageway

P2: Stairway

P3:Roof

P4: On a table

P5: Above a box

Figure 3.7 Installation locations of sensors used to detect vehicles and pedestrians

3.4.1 Data pre-processing

Data pre-processing is required because raw data contain noise such as randomised AMACs and AMACs from stationary devices such as printers. The raw data collected from the sensors were pre-processed as follows:

Filter 1: Delete randomised AMAC. MAC address randomisation is the increasing trend of device operating systems using a random, anonymous device identifier instead of the real MAC address when connecting to wireless networks. The goal of doing this is to increase user privacy by preventing network operators from being able track devices using the real MAC address as a consistent device identifier. Because we cannot track devices with randomised

AMACs, these should be removed. The data from the Wi-Fi packet sensor contains the organisationally unique identifier (OUI) of each AMAC, which can be used to determine whether the observed MAC address is randomised or not. Tables 3.3 and 3.4 summarise the raw observation data for vehicles and pedestrians, respectively. The rates of randomised AMAC addresses were 0.78 to 0.95, which means that only 5% to 22% of data were usable. As randomised addresses were not considered, the results may be under/overestimated.

Table 3.3 Raw vehicle observation data

Sensor	All records	Randomised AMAC	Real AMAC	Random rate	Time period
V1	35,457	33,487	1,970	0.94	12/5 8:00-18:00
V2	21,672	19,497	2,175	0.90	11/26 8:00-18:00
V3	5,536	4,761	775	0.86	11/26 8:00-18:00
V4	7,458	6,814	644	0.91	12/5 8:00 -18:00
V5	25,547	23,521	2,026	0.92	11/26 8:00-18:00

Table 3.4 Raw pedestrian observation data

Sensor	All records	Randomised AMAC	Real AMAC	Random rate	Time period
P1	5,182	4,178	1,004	0.81	2/6 8:00-18:00
P2	42,924	39,977	2,947	0.93	2/7 8:00-18:00
P3	25,303	19,664	5,639	0.78	2/7 8:00-18:00
P4	18,960	17,379	1,581	0.92	2/6 8:00-18:00
P5	92,281	87,801	4,480	0.95	2/7 8:00-18:00

Filter 2: Delete stationary devices. I deleted non-mobile devices because I am interested in moving devices. There may have been some Wi-Fi probe requests from non-mobile devices such as printers or laptops in offices. I defined non-mobile devices as those whose AMACs were observed over the course of 24 h per day, with the first observation time being between 00:00:00 AM to 00:05:00 AM and the last observation time being between 23:55:00 to 23:59:59. After this procedure, our AMAC data were ready to be analysed.

3.4.2 Vehicle detection analyses

I further aggregated the pre-processed data and counted vehicles from the video into 15 min periods. Many AMAC records were detected by sensor V1, which was installed at the refreshment corner located on the first floor of the Nursing Course Building, where many students come during lecture breaks. It may be possible to identify and remove these AMACs through RSSI. Because the students are nearer to the sensor than the vehicles, the signals from their devices are stronger, while those from vehicles are weak. Hence, I first defined a threshold RSSI to classify vehicles, then deleted AMACs with RSSIs stronger than this threshold. Because the distance between V1 and V5 was known and the time difference could be obtained from Wi-Fi packet sensor data, I matched the data from V1 and V5 and calculated the speed of the devices with matched AMACs. Generally, according to (Boehm et al., 2016), the speed of a cyclist is 20 km/h, so I categorised AMACs as originating from vehicles if their speed was greater than 20 km/h. The average RSSI of these AMACs was -88 dBm. I regarded AMACs as originating from within the building if the RSSI was greater than -88 dBm and deleted these from the V1 data set. Similarly, I also deleted AMACs that originated from inside the building from the V4 sensor data (located in the Animal Medical Centre). The road leading from V2 to V4 has turns and small bumps, so I manually observed the time needed to cycle between V2 and V4. Based on our observations, the average time was 151 s. Thus, I defined an AMAC as belonging to a vehicle if its time difference was less than 151 s. The average RSSI and median of the matched AMACs (V2 and V4) was -86 dBm. I removed AMACs from the V4 data set if their RSSI was stronger than -86 dBm. For sensor V4, due to a camera problem, I lost 3 sets of video data from 2018/12/5 12:30:00 pm to 2018/12/5 13:15:00, so I only had 37 samples for V4 and 40 data points for the other four sensors. I obtained an approximation function for each Wi-Fi packet sensor by comparing the Wi-Fi packet sensor data to the ground truth data. Because the person driving a car to university is mainly for commuting, thus I consider one unique AMAC address as one vehicle. Then a fitting function for the vehicle observation data was derived. The function of each sensor was selected by the higher R square value through comparing the value when fitting the data by linear fitting and exponential function fitting.

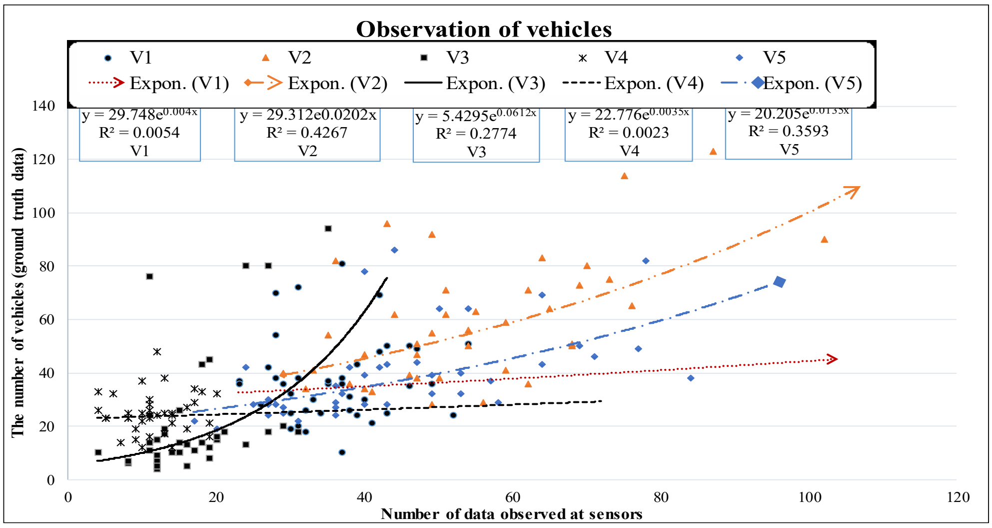

Figure 3.8 Estimating vehicle flow from Wi-Fi packet sensor data

From Figure 3.8, we can see that the relationship between the vehicle observation data collected from the Wi-Fi packet sensors and the ground truth data fit well to the exponential function for all five sensors. Moreover, the coefficients of determination were better for V2, V3, and V5, which were installed at the gates of the university. This may be because vehicles will slow down at the gate for entry permit inspection, which makes it easier to detect Wi-Fi packet signals from devices inside these vehicles. The large errors in V1 and V4 may be due to the obstruction of the building or signal attenuation because of distance. The coefficient of determination for V3 was much smaller than those for V2 and V5. Regarding the locations of these three sensors and the laboratory height test, V2 and V5 were on the table in the guard room, while V3 was on the ground because there is no guard room at the Nishi Bashi gate. The location of V3 was too low so the signal may have been blocked by vehicles, because the lower parts of vehicles are made of metal. In addition, the coefficient of determination for V5 was smaller than that of V2. This may be because the sensor types were different, although V2 and V5 were installed at similar heights. Based on these results, we can state that Wi-Fi packet sensors should be installed outside buildings and close to the road, and their heights should not be lower than vehicle windows. Although the correlation coefficients are not high, it can at least enlighten us there is a certain relationship between actual data and observed data, and further research (maybe a better filtering method) is needed.

3.4.3 Pedestrian detection analyses

I counted the number of pedestrians manually and used the results as our ground truth data. I set an imaginary cordon line near each sensor and counted pedestrians who crossed it. P3 was installed on the roof of a building, and the other four Wi-Fi packet sensors were installed at locations along roads traversed by pedestrians. I selected several time periods when there are generally more pedestrians for our analyses. The time periods varied from 15 min to 1 h. I filtered the Wi-Fi packet sensor data from the time periods for which we had ground truth data and regarded captured unique AMAC addresses as belonging to one person. Because vehicles cannot go into the detection area, the results could only be from devices carried by pedestrians. Then I derived a function to approximate the pedestrian traffic by comparing the Wi-Fi packet sensor data to the ground truth data.

From Figure 3.9, we can see that the relationship between the pedestrian observation data from the Wi-Fi packet sensors and the ground truth data fit best to a linear function, except in the case of P2. The slope of the curves when observing vehicles are gradually increasing, which means when the vehicles become density (Sometimes drivers have to queue through the gate one by one.), the detection rate of the Wi-Fi packet sensor will decrease. This may be caused by the obstruction of vehicles' bodies. However, it is rare for high-density students to pass by a

Wi-Fi packet sensor together. The location of P2 may be crowded than others, that's why P2 is not linear. Moreover, the coefficient of determination for P3 was very small. This may be because P3 was installed on the roof of a building, so many observations may have been lost due to obstruction by the building. In the cases of P2 and P3, insufficient data may have contributed to the difference between the results from these sensors and the others. There are several vending machines and a passageway near P4, but only the people who go into and out of the second canteen were counted in as the ground truth data, this is one of the factors that weaken the accuracy of the estimation function of P4. Based on these results, I confirmed that we can roughly estimate the flow of vehicles and pedestrians through Wi-Fi packet sensors, and the coefficients of determination were larger when observing pedestrians than when observing vehicles.

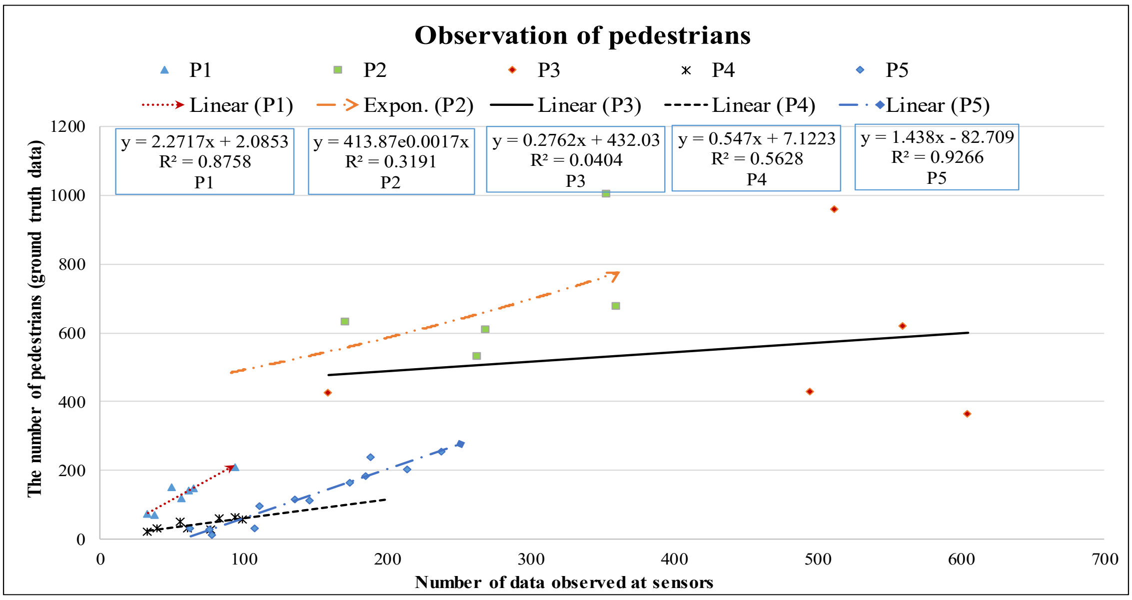

Figure 3.9 Estimating pedestrian flow from Wi-Fi packet sensor data

This suggests that Wi-Fi packet sensors may be more suitable for observing pedestrians. In addition, P1 and P5 have very good accuracy and can obtain a reliable number of flow, which inspires us that the Wi-Fi packet sensor could be a good way to accurately quantify the pedestrian flow volume.

3.5 Categorisation of observations by clustering analyses

Next, I investigated whether it could identify different types of users based on the observed data, but these may include observations that are not related to human movement. I analysed student behaviour while on campus. K-means clustering analyses (Syakur et al., 2018)were applied to each AMAC address to identify the type of device. From the pedestrian observation data collected from 2019/02/04 (Monday) to 2019/02/10 (Sunday), I deleted randomised AMACs and those devices that we only observed once. These data are summarised in Table 3.5.

Table 3.5 Data used for clustering analyses

All packages	1,207, 976
After deleting the random AMACs	133,231
AMACs were observed onc	1,161
Packages for analyses	132,070
AMAC number	2,853

In total, seven factors with 40 items were considered for K-means clustering analyses. These factors are described in Table 3.6.

Table 3.6 Factors considered for K-means clustering analyses

Factors	Definition
Observation ratio by sensor	For an AMAC, the number of observations by a sensor divided by the total observations of this AMAC 1 item per sensor installation point
Observation ratio by hour	Number of observations each hour divided by the total number of observations 24 items for each hour
Observation ratio by day of the week	Number of observations on a given day divided by the total number of observations 7 items, one for each day of the week
Observed days	Observed days of an AMAC
Number of capture sensors	Number of sensors captured by the same AMAC
Observation time	Time of an AMAC observation with respect to first and last observed time
Total observations	Number of observations of a particular AMAC

To use the K-means clustering method, the number of clusters, K, must be identified in advance. The elbow method was applied to determine the optimal K value. The results of the elbow method are shown in Figure 3.10. The vertical axis is the sum of the squares of the distances between samples and the centre of their cluster, and the horizontal axis represents the number of clusters. We can see that the optimal value of K was 3, so I used this value for our K-means clustering.

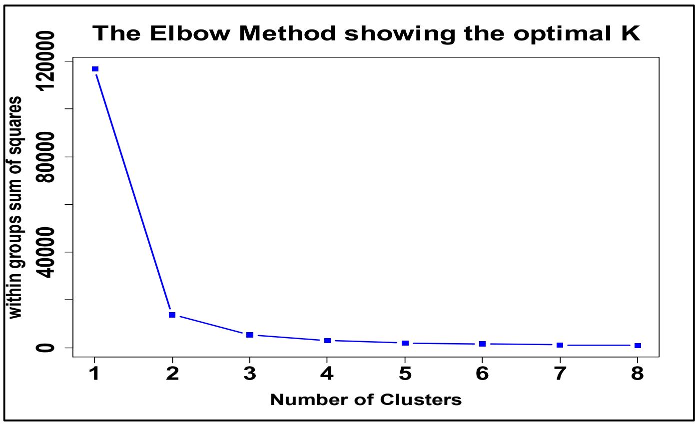

Figure 3.10 Results from the elbow method

The AMAC counts for each cluster are detailed in Table 3.7. We can see more than 96% of all AMACs were in cluster1, while cluster 2 and 3 only has a small part.

Table 3.7 AMAC count of each cluster

Cluster	**1**	**2**	**3**
AMAC counts	2,751	11	91

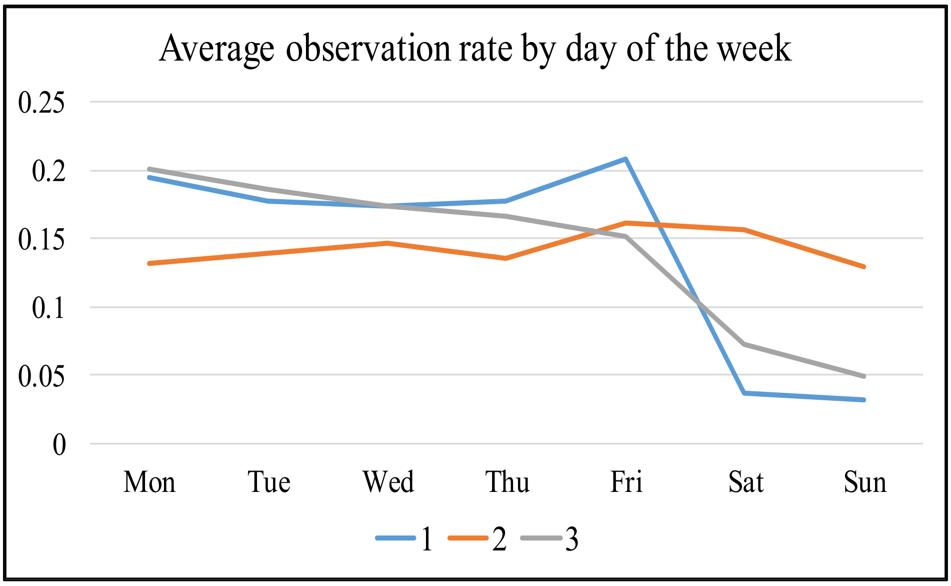

Figure 3.11 Average observation rate by day of the week

Figure 3.11 shows the average observation rate with respect to the day of the week. We can see that clusters 1 and 3 had a clear downward trend during the weekend, while there was no obvious change in cluster 2. In addition, cluster 3 had a larger observation rate than cluster 1 during the weekend.

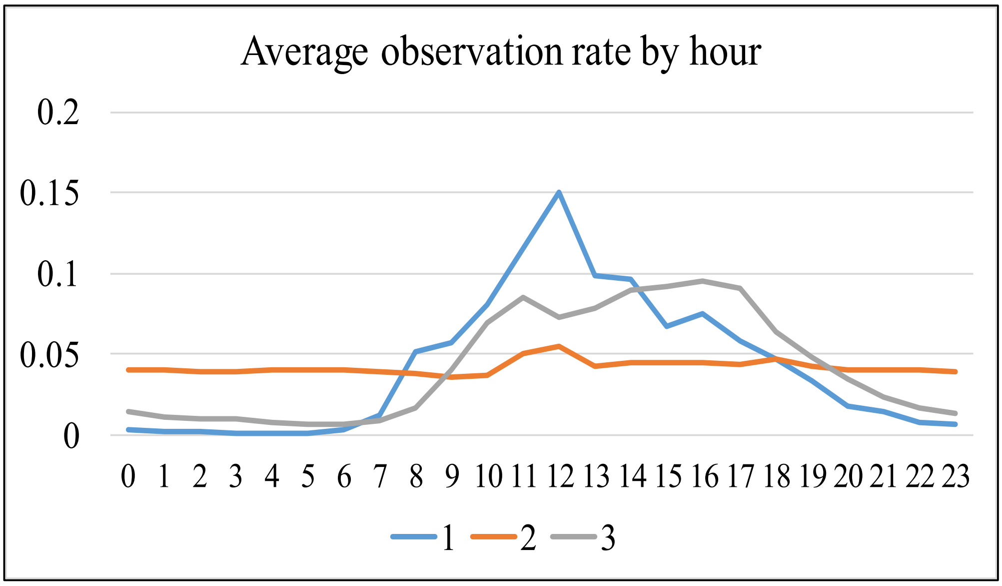

Figure 3.12 Average observation rate by hour

Figure 3.12 shows the average observation rate with respect to the hour. The rate for cluster 1 increased rapidly from 7 to 8 am, then began to fall from 12 onwards. In the case of cluster 3, the observation rate increased and decreased gently, then increased rapidly from 8 to 10 am. There were even some observations at midnight in cluster 3. Considering that the first class starts at 8:45 and morning classes end at 12:00, we can assume that devices in cluster 1 were most likely owned by undergraduate students; cluster 3 most likely represented graduate/research students because their schedule allows them to move more freely than undergraduates. Cluster 2 contained no obvious changes. The peak time of cluster 1 was 12:00, while cluster 3 had a peak at 11:00. This may because time constraints are stricter for undergraduates. Figure 3.13 shows the percentage of observations in each location per cluster. Devices in cluster 1 often appeared at the academic core of library. This is reasonable because examinations were held during the observation period. Cluster 2 appeared mostly at the entrance of the engineering building, which may be because sensor 3 was near the administration office of this faculty. There were also wireless printers and laptop computers in the administration office. Cluster 3 mainly appeared at the engineering building. Moreover, devices in cluster 3 were not often observed at the canteen. This may be because graduate students often buy lunch from the second floor of the first canteen building then eat at the laboratory. Although some choose to eat at the canteen, the first canteen is more popular than the second because it is more convenient. Hence, our observations are consistent with the habits of graduate students.

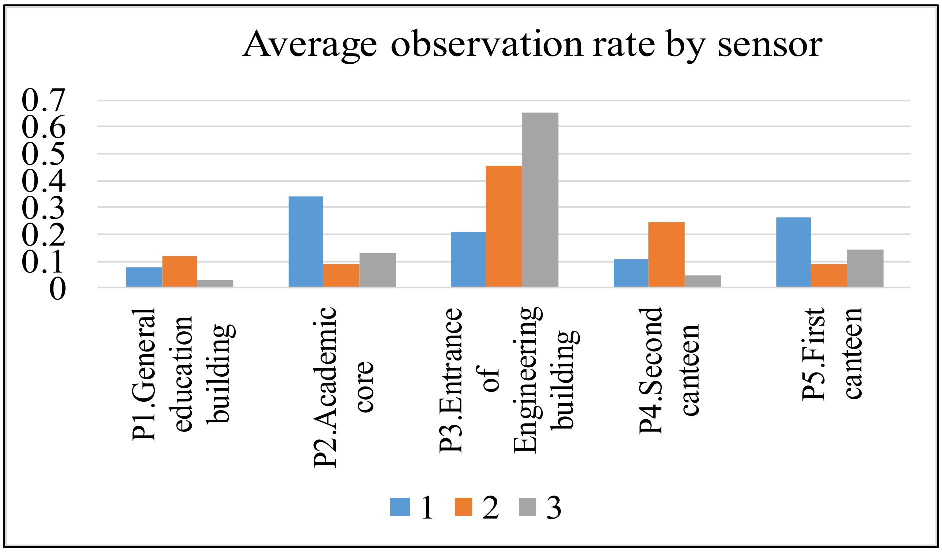

Figure 3.13 Percentage of observed locations

Table 3.8 Other characteristics per cluster

Cluster ID	# of Obs Days	# of Obs Hours	# of Obs Sensors	# of observations
1	2.50	51.55	2.31	23.16
2	6.64	154.68	1.55	2621.64
3	5.45	120.84	2.04	434.20

Table 3.8 shows other observation characteristics of each cluster. I use "#" to represent the average value of an indicator. The average observed hours in cluster 1 was 51.55 h; cluster 2 had the largest number of average observed hours, at 154.68 h; the average observed hours of cluster 3 was 120.84 h. These results are roughly consistent with the number of observed days. The average number of sensors at which each user was observed, reflects the areas of activity in each different cluster. We can see that users in clusters 1 and 3 often moved between more than two places, while those in cluster 2 moved between less than two places. Moreover, cluster 2 had a much higher number of observations for each AMAC.

Based on these analyses, I identified the characteristics of clusters, as summarised in Table 3.9. Characteristics of observations can be classified in a similar manner when sensors are installed in other places, such as city centres or tourist attractions.

Table 3.9 Results inferred from K-means clustering

Cluster ID	Tendency	Estimated attribute
1	Large sample size. Observed mainly in the morning. Few observations at night or weekends	Undergraduate students
2	Small sample size, observed continuously, 7 days a week, 24 h a day	Stationary devices
3	Observed mainly on weekdays but some observations at night. Often observed at the entrance of Eng. Bldg.	Graduate students

3.6 Spatial and temporal variation of pedestrian flow

In this section, I checked how the pedestrian flow of different locations changed over time. Figure 3.14 shows how AMAC counts were distributed among five Wi-Fi packet sensors (s1 to s5) between four time periods of a Monday (2019/02/04) when observing pedestrians. The size of the flow corresponds to the proportion of AMAC count it contains. The 'nonactive' group denotes the AMACs were observed in the previous time period but not observed in the present time period. It can be seen that a large part of the people became nonactive after lunchtime (t2), this may be because they left the campus in the afternoon before 17:00. It can also be seen that many people at the academic core (s2) and Engineering building (s3) tend to stay for several hours. Moreover, it can be seen that some people will change their study place such as the flow from t2_s2 to t3_s3 and from t2_s3 to t3_s2. The people who studied at the canteen can also be identified from t3_s5 to t4_s5.

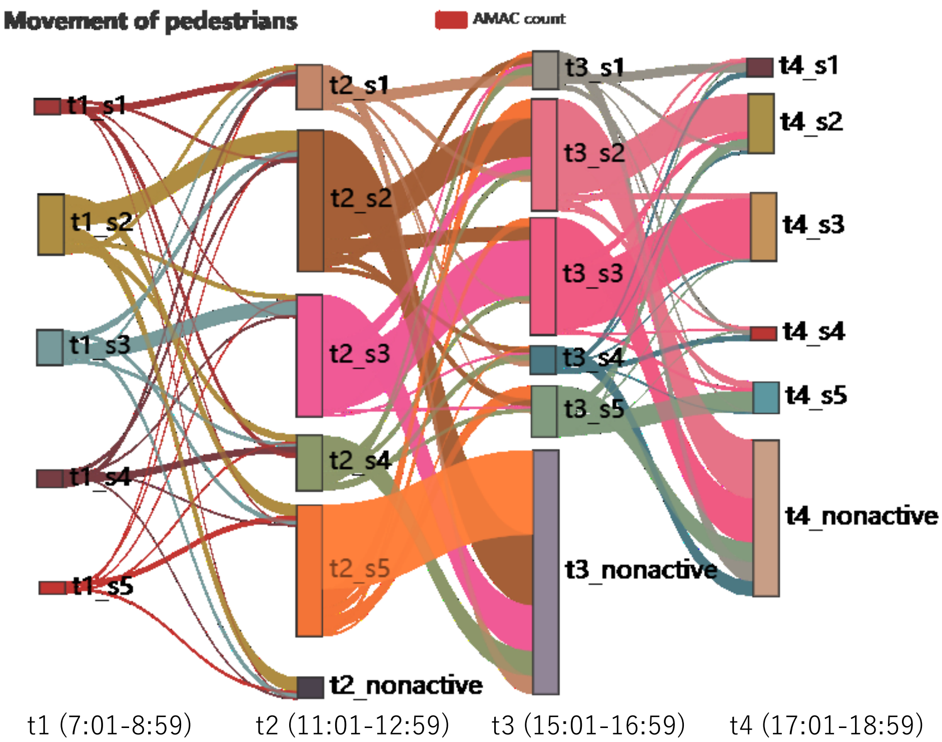

Figure 3.14 Spatial and temporal variation of pedestrian flow

3.7 Conclusion

This chapter explored how the environmental conditions of Wi-Fi packet sensors influence their detection capabilities. To this end, I carried out in-laboratory and on-campus experiments. Based on the data observed in our experiments, I investigated the factors influencing the detection capabilities of Wi-Fi packet sensors. It was found that the detection capacity varies between sensors and the detection rate varies with respect to sensor type and installation height. Outdoor sensors achieved better detection rates than indoor sensors. To collect more data, it is necessary to install Wi-Fi packet sensors outside buildings, and not near the floor and ceiling. Then I investigated whether it could estimate vehicle and pedestrian flows based on Wi-Fi packet sensor data. It was found that, when detecting vehicles, the relationship between the ground truth data and Wi-Fi packet sensor data fit well to the exponential function for all five sensors. Conversely, when detecting pedestrians, the relationship between the ground truth data and Wi-Fi packet sensor data can fit either a linear function or an exponential function. The coefficients of determination calculated when observing pedestrians were larger than those when vehicles were observed. This suggests that Wi-Fi packet sensors may be more suitable for observing pedestrians. Furthermore, it confirmed that we can cluster anonymous Wi-Fi packet sensor data based on movement trends and checked how Wi-Fi packet sensor data could reflect the spatial and temporal variation of pedestrian flow. This suggests that we can also study tourist behaviour using Wi-Fi packet sensors. For example, we can study the movement patterns of tourists by extracting their trajectories from data from multiple sensors. One of the main limitations of this chapter is that the findings in the in-laboratory experiment were not clearly reflected by the on-campus experiment. This should be considered in the future. It can increase the robustness of our results by increasing the number of observations and taking other factors, such as built environment and density of people, into account.

References

Andión, J., Navarro, J.M., López, G., Álvarez-Campana, M., Dueñas, J.C., 2018. Smart Behavioral Analytics over a Low-Cost IoT Wi-Fi Tracking Real Deployment. Wireless Communications and Mobile Computing 2018, e3136471. https://doi.org/10.1155/2018/3136471

Boehm, M., Ryeng, E., Haugen, T., 2016. Evaluating the Usage of Wi-Fi and Bluetooth Based Sensors for Pedestrian Counting in Urban Areas, in: Undefined. Presented at the European Transport Conference 2016Association for European Transport (AET).

Fukuda, D., Kobayashi, H., Nakanishi, W., Suga, Y., Sriroongvikrai, K., Choocharukul, K., 2017. Estimation of Paratransit Passenger Boarding/Alighting Locations Using Wi-Fi based Monitoring: Results of Field Testing in Krabi City, Thailand. Journal of the Eastern Asia Society for Transportation Studies 12, 2151–2169. https://doi.org/10.11175/easts.12.2151

Fukuzaki, Y., Mochizuki, M., Murao, K., Nishio, N., 2015. Statistical analysis of actual number of pedestrians for Wi-Fi packet-based pedestrian flow sensing, in: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, UbiComp/ISWC'15 Adjunct. Association for Computing Machinery, New York, NY, USA, pp. 1519–1526. https://doi.org/10.1145/2800835.2801623

Fukuzaki, Y., Mochizuki, M., Murao, K., Nishio, N., 2014. A pedestrian flow analysis system using Wi-Fi packet sensors to a real environment, in: Proceedings of the 2014 ACM International Joint Conference on

Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp '14 Adjunct. Association for Computing Machinery, New York, NY, USA, pp. 721–730. https://doi.org/10.1145/2638728.2641312

Kalogianni, E., Sileryte, R., Lam, M., Zhou, K., Van der Ham, M., Van der Spek, S.C., Verbree, E., 2015. Passive WiFi monitoring of the rhythm of the campus. Proceedings of The 18th AGILE International Conference on Geographic Information Science; Geographics Information Science as an Enabler of Smarter Cities and Communities, Lisboa (Portugal), June 9-14, 2015; Authors version.

Musa, A.B.M., Eriksson, J., 2012. Tracking unmodified smartphones using wi-fi monitors, in: Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, SenSys '12. Association for Computing Machinery, New York, NY, USA, pp. 281–294. https://doi.org/10.1145/2426656.2426685

Syakur, M.A., Khotimah, B.K., Rochman, E.M.S., Satoto, B.D., 2018. Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. IOP Conf. Ser.: Mater. Sci. Eng. 336, 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Xhafa, F., Leu, F.-Y., Hung, L.-L. (Eds.), 2017. Smart sensors networks: communication technologies and intelligent applications, Intelligent data centric systems. Academic Press, London.

Chapter 4: Estimation on real pedestrian count using Wi-Fi packet sensor

4.1 Introduction

In chapter 3 it was found that the observation result of the Wi-Fi packet sensor can be affected by several factors such as the setting height and sensor type, and a rough correlation exists between ground truth data and Wi-Fi packet sensor observations based on the one-sensor method. In this chapter, pedestrian flow is estimated utilizing two-sensors method considering installation conditions of the Wi-Fi packet sensors.

Counting the number of pedestrians is of great importance because of its usefulness in many fields. The pedestrian traffic volume information is important basic information for urban traffic planning which includes the analysis of walking speed, flow, density and space required by the pedestrians. If the pedestrian flow can be better understood, it will be possible to build statistics for pedestrian movements such as OD (Origin-Destination) pattern that can be used for developing and evaluating transportation plans. If crowding exists at a specific location at a specific time for example when there is an event such as a music festival or at a tourist area on holidays, it may cause conflict among people. Therefore, it is required for the management of pedestrian flow so as to ensure the safety of pedestrians by smoothening their movement.

In recent years, smartphones have become widespread. The abundant sensors and interfaces provide potential to researchers to study pedestrian behaviour through tracking the device they carry. For example, Versichele et al. (2012) analysed the complex spatiotemporal dynamics of visitor movements at a festival with 22 Bluetooth scanners. Yoshimura et al. (2017) analysed the differences between the pedestrian movement patterns on discount days and normal days in a shopping environment. Five Bluetooth sensors were installed in a historical centre of Barcelona to collect pedestrian data for one month. Daamen et al. (2016) compared data collection using counting cameras, Wi-Fi sensors and GPS trackers to estimate pedestrian traffic state during a large-scale event in an urban area. According to their study, the counting camera can give accurate local counts but have no speed and route information; the Wi-Fi data can get information on routes and travel times (speeds) but the penetration rate for Wi-Fi sensors was low; the GPS tracker data can provide detailed route choice information but not continuous, and the sample size is rather small compared to the Wi-Fi data. Alekseev and Lam (2019) compared the number of pedestrians estimated using Wi-Fi scanners and video records and developed a framework to calculate the number of pedestrians through conducting experiments on the university campus.

In comparison with video recording techniques, the new monitoring techniques can actively collect pedestrian data with larger spatial and time scales and can provide data in real-time. Ota et al. (2018) studied the acquisition rate of the Wi-Fi packet sensor considering the surrounding conditions, installation height, and the stay behaviour of the pedestrians of the sensor as well as the penetration rate of the devices with Wi-Fi turned ‘ON’. In their study, they set three levels of attenuation rate (level A=5%, level B=15%, level C= 25%) for the surrounding conditions, installation height and the approximate density of the pedestrians. For example, if the three factors are all level A for a sensor, its acquisition rate is 85.7%. Even there are some studies using automatic technologies to estimate pedestrian flow, studies using this kind of data to estimate pedestrian flow is still limited and they didn’t consider the effect of surrounding environment on the observations of the sensors. Therefore, the objective of this chapter is to quantify the influencing factors of the Wi-Fi packet sensor observations, such as height and surrounding conditions.

4.2 **Research area and data collection**

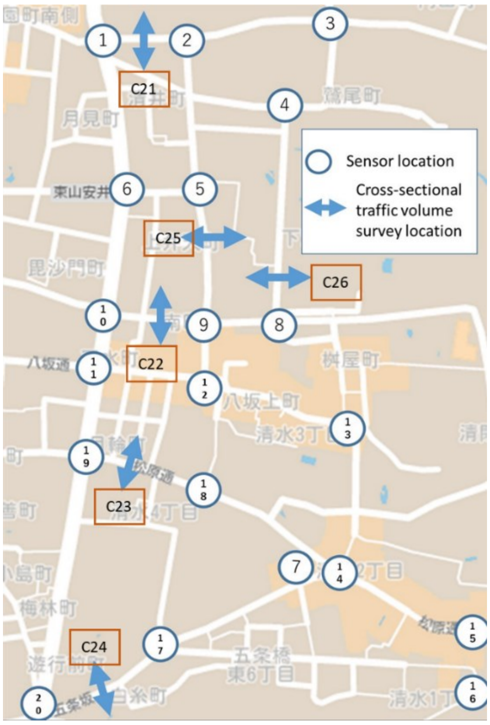

Figure 4.1 Data collection locations in Higashiyama area

The data used in this chapter were collected from the Higashiyama area around Kiyomizu Temple, which is one of the busiest tourist areas in Kyoto city, Japan. The detection area is about 0.6 km^2 (1,000 metres long and 600 metres wide). Figure 4.1 shows the study area on the map. 20 Wi-Fi packet sensors were equipped to collect data for 6 months (from 2017/10/1 to 2018/3/28). At the same time, a manual count survey was carried out to collect the real pedestrian flow data at six cross-sections (C21~C26) on 2017/11/12 (Sun) and 2017/11/13 (Mon).

4.2.1 The classification of the installation condition of the sensors

As is shown in the previous chapter, the different installation conditions of the Wi-Fi packet sensor will influence the observation counts. Based on the characteristics of the installation locations, three levels of the surrounding conditions and installation height are given as shown in Table 4.1.

Table 4.2 presents the sensors' installation information for six cross-sections where a manual count survey was carried out. C21 to C26 are the names of cross-sections. Sensor ID A and B represent the sensor identification numbers at both end of the cross-sections. Height A(B) and surrounding condition A(B) represent the height level and surrounding condition level of the sensor A(B).

Table 4.1 Installation conditions of sensors

Installation height	h_0	2 meters high or more, at a place of a good view that is not easily affected by the surrounding people.
	h_1	1.5m~2m, places can be seen from a person's height.
	h_2	70cm ~1m, places with poor visibility, blocked by people or vehicles.
Surrounding conditions	e_0	Well-reflected: inside a building that reflects radio waves easily (made by concrete or steel), or there is a building that reflects radio waves opposite to the sensor within 10 meters.
	e_1	Moderate reflection and absorption: there are no buildings within 10 meters that easily reflect radio waves, and there are few radio wave absorbers such as people and trees.
	e_2	Well-absorbed: there are no buildings that reflect radio waves opposite to the sensor but many radio waves absorbers such as people and trees.

Table 4.2 Installation information of sensors at cross-sections

Cross-section	**Sensor ID A**	**Sensor ID B**	**Height A**	**Height B**	**Surrounding condition A**	**Surrounding condition B**
C21	1	2	h_0	h_0	e_0	e_1
C22	9	10	h_2	h_0	e_2	e_1
C23	18	19	h_0	h_0	e_2	e_2
C24	17	20	h_2	h_1	e_0	e_1
C25	5	9	h_2	h_2	e_2	e_2
C26	4	8	h_1	h_2	e_2	e_2

4.3 **Estimation of pedestrian flow using two sensors method**

4.3.1 Relationship between the Wi-Fi packet sensor data and manual count survey data

The observation data were filtered before checking relationship with survey data since there are meaningless observations such as randomised AMACs. Firstly, the randomised AMACs are removed, then the AMACs observed by two sensors at the same time were removed based on RSSI and UNIXTIME. The AMACs with stronger RSSI and firstly observed were kept. Finally, the AMACs were observed only once were removed since they cannot be counted as a movement between two sensors.

The count survey recorded the pedestrians passed by the cross-section every 15 minutes. The AMACs observed by two sensors at the end of the count survey were also aggregated every 15 minutes. The dataset from 9:00 to 10:30 on Nov 13th at cross-section 21 were removed because the Wi-Fi sensors had no observation. By comparing the two datasets we can get the correlation function as Figure 4.2 shows. The intercepts of the six functions are set as 30 because here I have only six functions while I want to estimate five parameters. Unifying the intercept can help to reduce the number of parameters that need to be estimated. I checked the intercept of each function when their R^2 is largest and the intercept is between 19 to 59. Then I checked the R^2 of six functions and R^2 in Table 4.4 with intercept set as 30, 40 or 50 and found both R^2 of six functions and R^2 in Table 4.4 are larger than when set intercept as 40 or 50. Therefore here I set the intercept as 30.

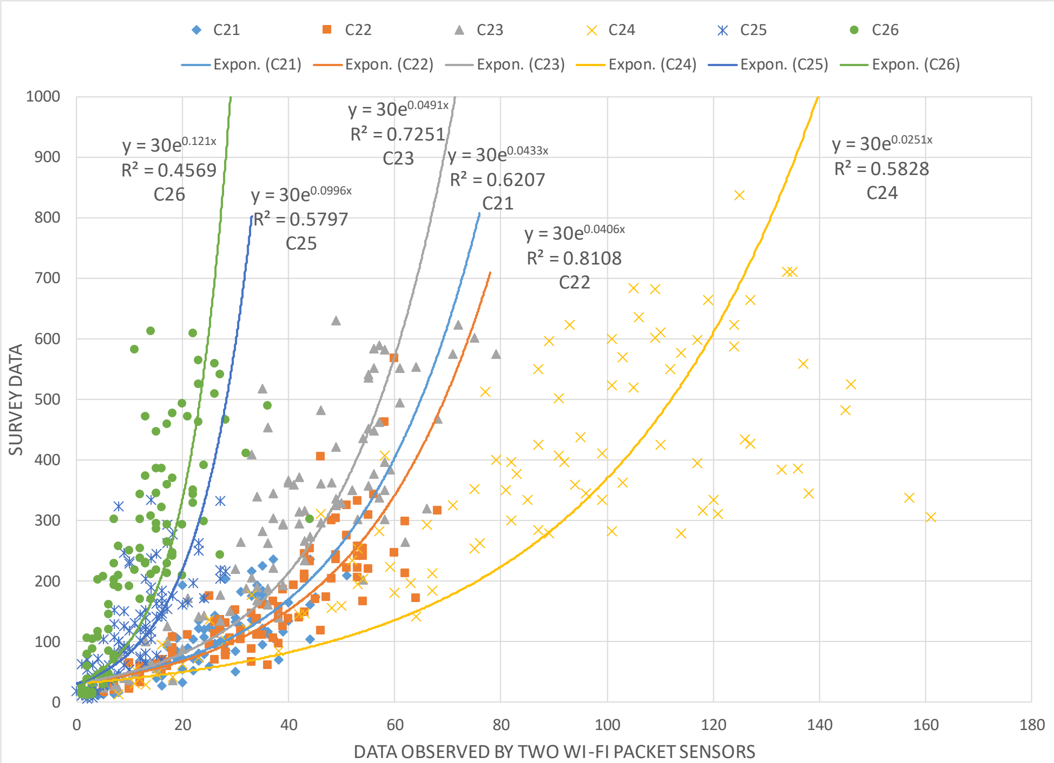

Figure 4.2 Correlation between Wi-Fi packet sensor data and survey data

The horizontal axis is AMAC count data from the Wi-Fi packet sensor and the vertical axis is the pedestrian count obtained by manual count survey. It can be found that the correlation may fit well with the exponential function (the detection ability may decrease gradually when the number of pedestrians increases). This is reasonable because when there are too many people, the body of people can be regarded as an obstacle and will absorb the Wi-Fi signal. Furthermore, the unique AMACs under the same pedestrian traffic volume differs greatly for the six cross-sections. The acquisition rate was quite low at cross-sections C26 and C25 while it was high at C24. From the figure 4.1, C26 and C25 are located on Nene-no-michi and Shimogawara dori that are very popular with tourists. As mentioned above, the Wi-Fi radio waves can be absorbed by human bodies, and the acquisition rate tends to decrease as the pedestrian traffic volume increases. Therefore, it can be said that the number of observed unique AMACs may decrease with respect to the pedestrian volume at these two locations. Another possible reason is that there are streets between C25 and C26 such as Ishibe-koji, which are also well visited by tourists, and the people who use these streets cannot be observed by the two sensors leading to less matched AMACs. On the other hand, C24 is located at a rather wide road and there are sightseeing buses passing by; there may have some observations from the tourists inside buses by sensors 20 and 17. Therefore, the number of observed unique AMACs seems to increase.

4.3.2 Quantifying factors influencing the observation of Wi-Fi packet sensor

Based on Figure 4.2 we can see the correlation between survey data and Wi-Fi packet sensor data were suitably fitted with an exponential function. As mentioned above, Ota et al. (2018) studied the acquisition rate of the Wi-Fi packet sensor considering the surrounding conditions and installation height. However, in their study, they just simply assume the three levels of attenuation rate (level A=5%, level B=15%, level C= 25%) for the surrounding conditions and installation height. If we want to analyse pedestrian flow quantitatively, it will be necessary to estimate the acquisition rate of each sensor so that the Wi-Fi packet observations can be expanded to the real pedestrian flow. If the relationship between the pedestrian traffic volume at cross-sections and the installation conditions of the sensors can be understood clearly, it will be possible to estimate the pedestrian volume more accurately.

Since there are only six cross-sections in this survey, the number of parameters that can be estimated is up to six. Therefore, we decide to estimate the coefficient of the regression curve shown by Figure 4.2 with the installation height (h_0, h_1) and surrounding conditions (e_0, e_1) as the explanatory factors. The equation between the ground truth data and Wi-Fi packet sensor data is assumed as:

$$y = 30e^{\beta x} \tag{4.1}$$

y: actual pedestrian flow volume, x:number of AMACs observed by two sensors.

Here, β is the result of the combined effect of the installation height and the surrounding conditions of the Wi-Fi packet sensor. Let:

$$\beta = c * h_0^{x_{h_0}} * h_1^{x_{h_1}} * e_0^{x_{e_0}} * e_1^{x_{e_1}} \tag{4.2}$$

Definitions of variables in the model is shown in Table 4.3:

Table 4.3 Definitions of variables

c	Constant, when height is h_2 and surrounding condition is e_2 (unknown variable)
h_0, h_1	Explanation of y when set height is h_0 or h_1 (unknown variable)
e_0, e_1	Explanation of y when surrounding condition is e_0 or e_1, (unknown variable)
x_{h_0}, x_{h_1}	Number of sensors of the link when set height is h_0 or h_1 (it should be 0, 1 or 2)
x_{e_0}, x_{e_1}	Number of sensors of the link when surrounding conditions is e_0 or e_1 (it should be 0, 1 or 2)

The least squares method was applied to calculate β of the trend lines in the Figure 4.2 and also be used to estimate the parameters (h_0, h_1, e_0, e_1) of the equation (4.2). Equation (4.2) can be transferred to a linear equation by taking the logarithm of both sides. Table 4.4 shows the estimation result.

Table 4.4 Parameter estimation result

Parameter	Value	T-stat	R^2
c	0.09960	-3.19	0.70
h_0	0.84135	-0.36	
h_1	0.84605	-0.17	
e_0	0.88279	-0.12	
e_1	0.48450	-0.84	

The larger estimates of parameters will result in a larger rise of the curve, meaning that the sensor observations become less when the pedestrian volume is the same, in other words, the acquisition rate of the Wi-Fi packet sensor becomes lower. From this point of view, the h_0 is smaller than h_1 in Table 4.3, meaning that when putting the Wi-Fi packet sensors at a higher place (2 meters), the observation efficiency will be better. Maybe it is because it can reduce the obstruction of the human body to the Wi-Fi probe. e_0 is larger than e_1 meaning that when the Wi-Fi packet sensors were installed inside a building or the surrounding conditions are easy to reflect the Wi-Fi probe, the acquisition rate is lower. Considering that the coefficient can be interpreted as 1 in the case the installation height and surrounding conditions are h_2 and e_2. The acquisition rate is higher if the parameter has a lower value. In conclusion, h_0 is the most efficient in terms of installation height, and e_1 is the most efficient considering the effect of radio wave absorption and reflection. Moreover, e_0 is less than 1 meaning that the absorption has a more obvious influence on the observation of Wi-Fi packet sensor compared with reflection. R square is 0.7 means the parameters can explain 70% of the variation of β. The t values for the parameters seem not statistically significant (the absolute value is less than 2). This may be because the sample size is too small and more sample is needed to improve the estimation.

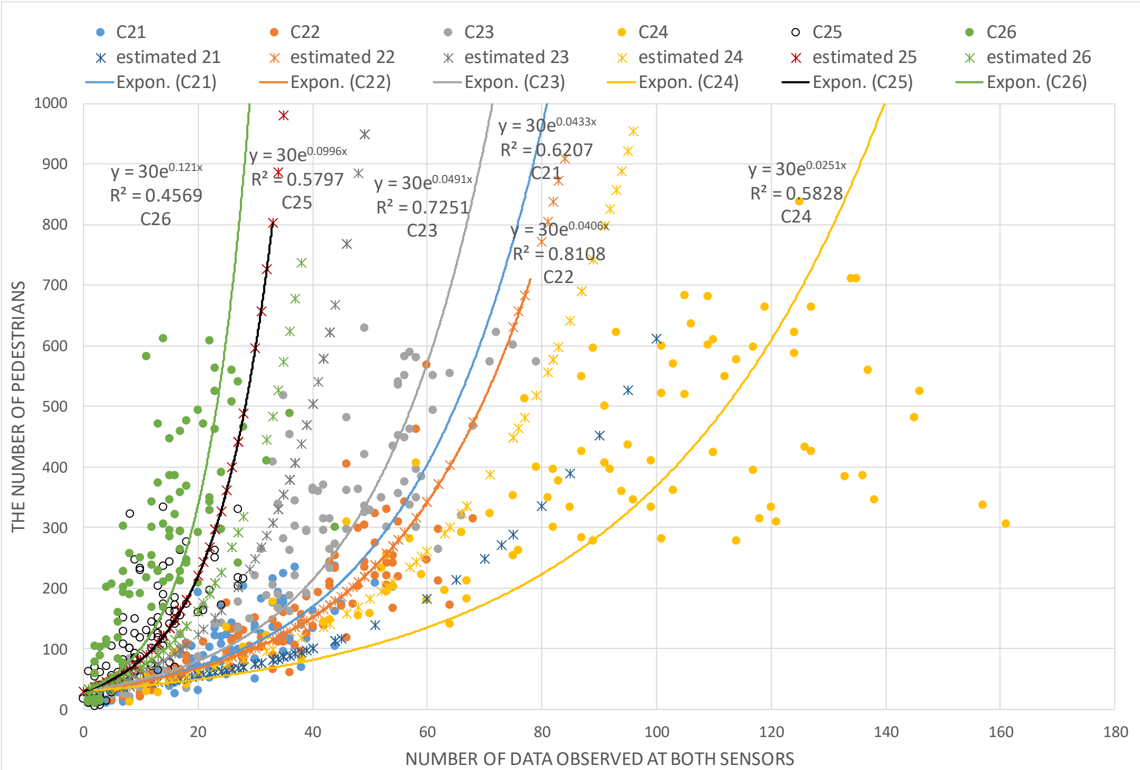

Figure 4.3 Estimated result based on the parameters

Figure 4.3 compared the estimated pedestrian flow based on the parameters in Table 4.4 with the trend line of each cross-section. It can be seen that, for cross-sections C22 and C25, the model can well estimate the pedestrian flow based on the Wi-Fi packet sensor observations. However, the models are underestimating for C21 and C26, and overestimating for C23 and C24**.** Since we only counted pedestrians at 6 cross-sections, we cannot further improve the model. It is necessary to increase the observation cross-sections and verify parameters in the future. It is also necessary to verify the reliability of the evaluation levels of the installation conditions in Table 4.2.

4.4 **Conclusions**

In this chapter, I studied how to estimate the real pedestrian flow volume based on the Wi-Fi packet sensor observations. A two-sensor method was used to filter out the unique MAC address data that belong to the pedestrians passing by. Through comparing with the manual count ground truth data, it can be confirmed that there has a strong correlation between pedestrian traffic counts at a cross-section and the number of unique AMACs matched by two sensors. In addition, it is revealed appropriate to use an exponential function as a regression curve. Next, I examined whether the coefficient of the exponential function could be explained by the sensor installation conditions. I then quantified the influence factors with data collected at six cross-sections in the Higashiyama area of Kyoto city. As a result, if the sensor installation conditions can be known, it will be possible to estimate the real pedestrian flow of a segment.

The result of this chapter can therefore help study pedestrian flow characteristics such as the relationship between speed, flow and density and their fundamental diagrams on walkways using Wi-Fi sensor data. As an essential transportation mode, walking has been promoted by the government of more and more countries for broader sustainability and economic development strategies (Mp, 2017; Salvo et al., 2021; Guariguata et al., 2021; United Nations, 2021; Hirst and Dempsey, 2022). For cities to enhance walking activity, it is important to estimate how the walking trips are likely to distribute along the area in time and space. The method in this chapter can help understand pedestrian walking behaviour under different circumstances, from daily trips, movement on holidays, and special events. Since only 6 cross-sections were observed this time, there remains a problem in the reliability of the parameters. It is considered necessary to continue the study by increasing the number of cross-sections in the future.

References

Alekseev, N., Lam, W., 2019. Estimation of Pedestrian Flow Based on Wi-Fi Data and Video Cameras. Journal of the Eastern Asia Society for Transportation Studies 13. https://doi.org/10.11175/EASTS.13.93

Daamen, W., Yuan, Y., Duives, D.C., Hoogendoorn, S.P., 2016. Comparing three types of real-time data collection techniques: Counting cameras, Wi-Fi sensors and GPS trackers. Proceedings of Pedestrian and Evacuation Dynamics 2016.

Guariguata, L., Unwin, N., Garcia, L., Woodcock, J., Samuels, T.A., Guell, C., 2021. Systems science for developing policy to improve physical activity, the Caribbean. Bulletin of the World Health Organization 99, 722–729. https://doi.org/10.2471/BLT.20.285297

Hirst, D., Dempsey, N., 2022. Active travel: Trends, policy and funding.

Mp, A.J., 2017. Cycling and Walking Investment Strategy 38.

Ota, K., Omura, M., Tsujido, F., Asao, K., Nishida, J., 2018. Wi-Fi 歩行者流動センサによる計測値 からの 実数推定手法. 第 58 回土木計画学研究発表会・講演集 02–07.

Salvo, D., Garcia, L., Reis, R.S., Stankov, I., Goel, R., Schipperijn, J., Hallal, P.C., Ding, D., Pratt, M., 2021. Physical Activity Promotion and the United Nations Sustainable Development Goals: Building Synergies to Maximize Impact. Journal of Physical Activity and Health 18, 1163–1180. https://doi.org/10.1123/jpah.2021-0413

United Nations, 2021. Sustainable transport, sustainable development. Interagency report for second Global Sustainable Transport Conference.

Versichele, M., Neutens, T., Delafontaine, M., Van de Weghe, N., 2012. The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. Applied Geography 32, 208–220. https://doi.org/10.1016/j.apgeog.2011.05.011

Yoshimura, Y., Amini, A., Sobolevsky, S., Blat, J., Ratti, C., 2017. Analysis of pedestrian behaviors through non-invasive Bluetooth monitoring. Applied Geography 81, 43–51. https://doi.org/10.1016/j.apgeog.2017.02.002

Chapter 5: Pedestrian level of service (PLoS) measurement based on Wi-Fi packet sensor data

5.1 **Introduction**

Massive visitors bring not only economic benefits but also impacts and pressures to tourist attractions. The residents may suffer from heavy congestion when driving in these areas because of the crossing of pedestrians and queuing at bus stations. The visitors suffer from crowding caused by too many people or even security threat to their sightseeing experience. In order to relax such problems, it is necessary to understand the travel demand of visitors and evaluate pedestrian performance (whether the pedestrians can move smoothly) of that area so that the authority or community can manage and control movement of people. According to the research of (McKinsey & Company, & World Travel & Tourism Council, 2017), destinations can mitigate overcrowding by adopting the right mix of tactics which include smoothening visitors over time and spreading visitors across sites, adjusting pricing to balance supply and demand, regulating accommodation supply and limiting access and activities.

Based on chapter 4, it is possible to roughly estimate the real pedestrian count of a street utilizing Wi-Fi packet sensor data by referring to the installation conditions of the sensor. This chapter then tries to evaluate a crowding level in tourist areas with the concept of pedestrian level of service (PLoS). The concept of LoS (Level of Service) is proposed from the Highway Capacity Manual (HCM) and can be used for the assessment of the operational performance of a roadway facility. The LoS is intended to represent user-perceived quality of service and can be defined as a measurement of operational conditions within a traffic stream, generally in terms of travel time, speed, freedom to manoeuvre, traffic interruptions, comfort and convenience (HCM, 2000). Among various LoS models, the most common measure in evaluating the quality of walking conditions of a street for pedestrians is using the pedestrian level of service (PLoS). The PLoS can be defined as "*an overall measure of walking condition on route, path and facility and it reflects users' perception in terms of sense of mobility, comfort and safety*" (Gallin, 2001). The development of PLoS measures has received considerable attentions from academic researchers (Mōri and Tsukaguchi, 1987; Zhao et al., 2014; Kadali and Vedagiri, 2015; Nowar

Raad and Burke, 2017; Banerjee et al., 2018; Rahul and Manoj, 2020; Ujjwal and Bandyopadhyaya, 2021; Molyneaux et al., 2021) and practitioners (Otak, 1997; Croft et al., 2013; Transportation Research Board, 2016; AASHTO, 2021), as society seeks to improve the level of pedestrian activity to increase the share of walking as a mode of transport that may contribute to reduce car usage, to lessen the traffic congestion and associated environmental impacts. Initially, the traffic engineers assessed PLoS using methods similar to LoS assessment methods for traffic facilities, i.e. based on pedestrian flow volumes and capacity of the pedestrian facility (Fruin, 1971; Mōri and Tsukaguchi, 1987). Later, researchers tried to integrate qualitative factors into the assessment of LoS offered by pedestrian facilities which marked an important advance in the field of PLoS assessment (Sarkar, 1993; Khisty, 1994; Parida et al., 2007). Ghani et al. (2015) proposed a method to audit pedestrian infrastructure in a heritage site using Pedestrian Index (P-Index) method. The method is based on a star rating system including four indicators namely mobility, safety, facility and accessibility. Further, researchers have used other quantitative factors such as footpath width, shoulder width, buffer zone width and presence of on-street parking for PLoS assessment. Table 5.1 presents a list of different methodological approaches and factors considered by various pedestrian studies (Gr et al., 2018).

Various studies have focused on PLoS model development at intersections (Marisamynathan and Vedagiri, 2019), sidewalks (Tan et al., 2007; Gr et al., 2018), midblocks (Kadali and Vedagiri, 2015), stairways (Wen et al., 2012), and roadway segments (Asadi-Shekari et al., 2013a). Ujjwal and Bandyopadhyaya (2021) developed a comprehensive PLoS assessment model for mixed land-use of urban areas (having residential, commercial or shopping, and office activity in the same place). Based on the face-to-face interview survey they collected the importance rating score of the 24 walking encouragement and discouragement factors from 550 pedestrians. Then the PCA (Principal Component Analysis) was conducted to define the most important influencing factors. Finally, six important parameters were identified for PLoS assessment including safety issues under pedestrian traffic interaction, pedestrian convenience and sense of security, pedestrian walking comfort and so on. Ahmed et al. (2021) introduced a new pedestrian crossing level of service method to promote safe crossing in urban areas. According to Nowar Raad and Burke (2017), Gr et al. (2018) and Ahmed et al. (2021), the main conventional techniques for data collection in PLoS studies are direct observation (Anciaes and Jones, 2018), video techniques (Teknomo et al., 2000; Asadi-Shekari et al., 2013b; Al-Mukaram and Musa, 2020), and questionnaires (Zhang and Prevedouros, 2003; Zahid et al., 2020; He et al., 2020). Simulation methods, regression analysis, and point systems are the main analytical methods that are used to rate the street's condition. However, using direct observation solely can generate biased results, because it is purely dependent on the researcher's perception. Similarly, using only a questionnaire method of data collection limits the results to the respondent's perception. Besides, the questionnaire survey method cannot reflect the situation in real-time and cost more and investigators. As a consequence, there is no universal method that can be used to evaluate PLoS everywhere because it should consider the different characteristics of each place. On the other hand, the Wi-Fi packet sensor can be used to collect pedestrian count data without the necessity of the researcher's and respondent's cooperation therefore can acquire data of crowds objectively. Moreover, it is possible to acquire long-term and continuous observation data, therefore the Wi-Fi packet sensor data can be an ideal data source for the evaluation of PLoS.

Table 5.1 Summary of PLoS studies (Gr et al., 2018)

Authors (Year)	Methods	Factors considered
Sarkar (1993)	Scoring System	Convenience, comfort, safety, continuity, system coherence and attractiveness.
Khisty (1994)	Scoring System	Comfort, convenience, continuity, attractiveness, system coherence, safety, security.
Dixon (1996)	Scoring System	Path conflicts, amenities, motor vehicle LOS, maintenance problems, provision of basic facilities and provision of multiple modes.
Landis et al. (2001)	Ordinary Least Regression	Lateral separation factors, traffic volume, speed of the vehicle, driveway access frequency and volume
Gallin (2001)	Scoring System	Sidewalk width, sidewalk surface, comfort, walk environment, potential for vehicle conflict, crossing facilities and pedestrian volume.
Muraleetharan et al. (2005)	Ordinary Least Regression	Sidewalk width and separation, pedestrian volume, flow rate and bicycle events
Parida et al. (2007)	Scoring System	Footpath width, footpath surface, continuity, comfort, safety, encroachment, potential to vehicle conflict, crossing facilities, walking environment and pedestrian volume
Asadi-Shekari et al. (2014)	Scoring system	Footpath surface, footpath, corner island, width of footpath, tactile pavement(guiding), tactile pavement(warning), signal, seating area, drinking fountain, buffer, traffic lanes, crossing, facilities, furniture
Talavera-Garcia and Soria-Lara (2015)	Scoring system	Sidewalk width, sidewalk surface, walking distance
Aghaabbasi et al. (2016)	Scoring system	Sidewalk width, sidewalk surface, ramps, tactile pavements, utilities and landscape

Overall, the main aim of this chapter is to develop a quantitative PLoS measurement method appropriate for the tourism context that reflects the perceived comfort or safety of tourists in Higashiyama district based on the Wi-Fi sensing data. This method only needs the sidewalk width data to derive the pedestrian flow rate and PLoS level. This study might be considered as a guideline for evaluating pedestrian level of services for such tourist areas or other business areas.

5.2 Data preparation

The Wi-Fi packet sensor observation data used in this chapter is collected by 20 sensors in Higashiyama area, same as chapter 4. Figure 5.1 shows the study area on the map. 20 Wi-Fi packet sensors were equipped to collect data for 6 months (from 2017/10/1 to 2018/3/28). Several famous point of interest (Yasaka Shrine, Entoku Temple, Kodai Temple and Kiyomizu Temple) are circled and pointed out with arrows. Table 5.2 records the installation conditions of all Wi-Fi packet sensors. The $h_0 \sim h_2$ in the height column and $e_0 \sim e_2$ in the surrounding conditions column represent different levels. The detailed explanation of these levels can be found in chapter 4.

Figure 5.1 Coverage of observation locations in Higashiyama.

Table 5.2 Installation conditions of sensors

Sensor ID	Height	Surrounding conditions	Sensor ID	Height	Surrounding conditions
1	h_0	e_0	11	h_0	e_1
2	h_0	e_1	12	h_2	e_2
3	h_2	e_2	13	h_0	e_2
4	h_1	e_2	14	h_0	e_1
5	h_2	e_2	15	h_2	e_2
6	h_0	e_1	16	h_0	e_1
7	h_0	e_1	17	h_2	e_0
8	h_2	e_2	18	h_0	e_2
9	h_2	e_2	19	h_0	e_2
10	h_0	e_1	20	h_1	e_1

5.3 **Criteria for evaluating PLoS**

According to (HCM, 2000), the PLoS rating of a walkway is determined on the basis of two decision variables: the average pedestrian space and the pedestrian flow rate. Based on the study in chapter 4, it is possible to estimate the real pedestrian flow count of a road segment using Wi-Fi packet sensor data. The criteria of PLoS based on pedestrian flow rate is therefore used to evaluate the pedestrian flow performance in Higashiyama area. Table 5.3 shows the six-level scale PLoS criteria developed by Highway Capacity Manual (HCM, 2000) .

Table 5.3 Pedestrian walkway LOS (adapted from the HCM 2000)

Pedestrian LOS	Flow rate (ped/min/m)	Characteristics
A	<= 16	Free speed, no conflict.
B	16-23	Free speed, respond to other pedestrians.
C	23-33	Normal speed, reverse-direction or crossing movements can cause minor conflicts.
D	33-49	Restricted to select walking speed freely, high probability of conflict of crossing or reverse flow movements, reasonably fluid flow but friction and interaction between pedestrians is likely.
E	49-75	Virtually all pedestrians restrict normal walking speed, volumes approach the limit of walkway capacity, with interruptions to flow.
F	Flow rate varies	All walking speeds are severely restricted, and forward progress is made only by shuffling. Frequent, unavoidable contact with other pedestrians.

Level A represents the best condition scenario in terms of the road passability, level F represents the worst condition scenario (i.e., very congested/unsafe/uncomfortable). If the streets are too congested, people may feel uncomfortable and unsafe to use the street which contributes to the feeling of stress (Krupat, 1985). According to (Shamsuddin and Ujang, 2008), the presence of people can increase their feeling of safety in using the street. If there are too many people it will become an unsafe environment. Street users will avoid using streets that are too congested. However, the feeling of crowding is different for each of users with different purposes. The tourists who walk mainly for pleasure may feel satisfied and positive emotions while commuters running for catching a bus will feel negative emotions. There are also some positive effects of crowding. For example, in tourist attraction sites, crowding makes a street lively and inviting. The crowds also reflect vibrancy. People walking along the street during crowded situation tend to walk much slower for shopping purpose. This was because the shoppers tend to stroll and stop to look in windows (Al-Azzawi, 2004). According to (Radisya Pratiwi et al., 2015), during a crowded situation (festival), the pedestrians could observe the environment in detail because they walked at a low speed and they tend to value the availability of amenities the most. Therefore, it is best to keep the pedestrian flow neither too much nor too less. Maybe the levels C and D are the ideal PLoS levels for tourists to have a good experience.

5.4 **Analysis of the PLoS under different scenarios**

Firstly, I checked the tendency of the AMAC counts during the observation period. Figure 5.2 shows the AMAC count distribution of the whole observation period. The horizontal axis represents the date and the vertical axis represents the number of observed AMACs. The date is in red- orange when it is a weekend or holiday, and we can see more AMACs were observed on weekends and holidays than on weekdays. Moreover, it can be seen that there was overall the largest number of visitors in November 2017 and the least number of visitors in January 2018. This may be because November is the best time (mid-October to mid-December) for viewing maples in Kyoto.

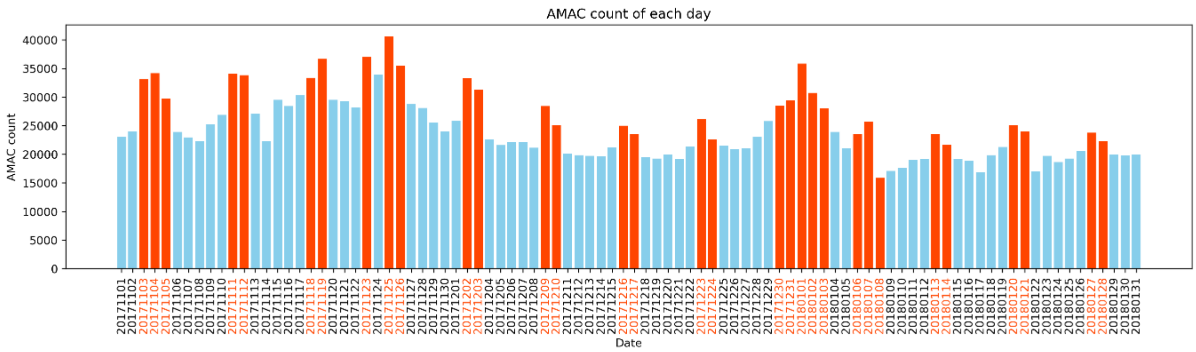

Figure 5.2 AMAC count distribution

5.4.1 Comparing PLoS on weekdays and weekends

Firstly, the PLoS on weekdays and weekends (holidays) were compared because there was an obvious difference between the AMAC observations on weekdays and weekends (holidays). The pedestrian flow rate is the total number of pedestrians crossing the given cross-section divided by the analysis period and the sideway width. The width of a sideway is collected through a field survey.

Figure 5.3 shows the name and location of links on the map. Based on the environment of each link I classified the links into 4 types as shown in Table 5.4. Figures 5.4-5.7 show the environment of a sample link of each type. Table 5.5 gives the name and location of each link. Sensor A/B on behalf of the sensor ID at the end of a link.

Figure 5.3 Link names and locations

Table 5.4 Classification of links.

Type	Link name	Characterisation
1	L1-L5	Main road for vehicles
2	L6, L7, L8, L9, L13, L14, L16, L17, L18	Good for walking, less POIs
3	L10, L15, L19, L22, L23, L24, L25	Very good for walking, many traditional buildings and POIs
4	L11, L20, L26, L12, L21, L27, L28	Heading to Kiyomizu Temple, many POIs, very popular

Figure 5.4 Street view of L1-L5

Figure 5.5 Street view of L18

Figure 5.6 Street view of L22

Figure 5.7 Street view of L12

Table 5.5 Link name and locations

Link name	Sensor A	Sensor B	Link name	Sensor A	Sensor B	Link name	Sensor A	Sensor B	Link name	Sensor A	Sensor B
L1	1	6	L8	5	6	L15	9	12	L22	3	4
L2	6	10	L9	9	10	L16	12	18	L23	4	8
L3	10	11	L10	11	12	L17	2	3	L24	8	13
L4	11	19	L11	19	18	L18	8	9	L25	13	14
L5	19	20	L12	17	20	L19	12	13	L26	14	15
L6	1	2	L13	2	5	L20	14	18	L27	16	17
L7	1	4	L14	5	9	L21	7	17	L28	15	16

Figures 5.8 to 5.15 shows the estimated PLoS distribution of different links. The vertical axis represents time period and the horizontal axis represents the name of links. The different colour shows the different PLoS level. Through comparing the PLoS distribution on weekdays (Figures 5.8 to 5.10) and on weekends or holidays (Figures 5.11 to 5.15), it can be recognised that the PLoS on weekends or holidays ranks higher over a longer time and more links than that

of weekdays. Moreover, the links L12, L28, L11 have a higher PLoS ranking than other places during some time period on weekdays. The links L12, L28, L11, L18 and L22 have extremely high PLoS levels on weekends and holidays. These links are consistent with the frequent patterns identified in chapter 6. The links that have obviously high PLoS level (L11, L28, L12), mostly belong to group 4 in Table 5.4 followed by group 3 (L22) and group 2 (L18). This means the Kiyomizu temple plays a key role in attracting tourists and visitors prefer narrower streets with fewer vehicles, which are closer to traditional buildings and POIs and are more suitable for walking tour. Among these links, the links L12 and L28 start to have high PLoS levels in the morning around 10:00, while the time period when links L11, L18 and L22 have high PLoS levels happens mainly in the afternoon. The reason may be because there are tourist buses going to the parking place at sensor 7 by using link L12. The parking place near sensor 7 is a special sightseeing parking lot with a large area dedicated to tourist buses and taxis. The vehicles can only access this parking lot through links L12 and L21. Some observations may come from the tourist on the bus and some of the tourist buses come rather early. Figure 5.10 shows the PLoS of a weekday of neither during the popular maple season (mid-October to mid-December) nor the special holidays. It can be seen that most of the roads were not crowded at all except the link L11 which is along the 'Golden route' identified in chapter 6. As a consequence it can be said that the PLoS based on the Wi-Fi packet sensor data can reflect pedestrian flow performance.

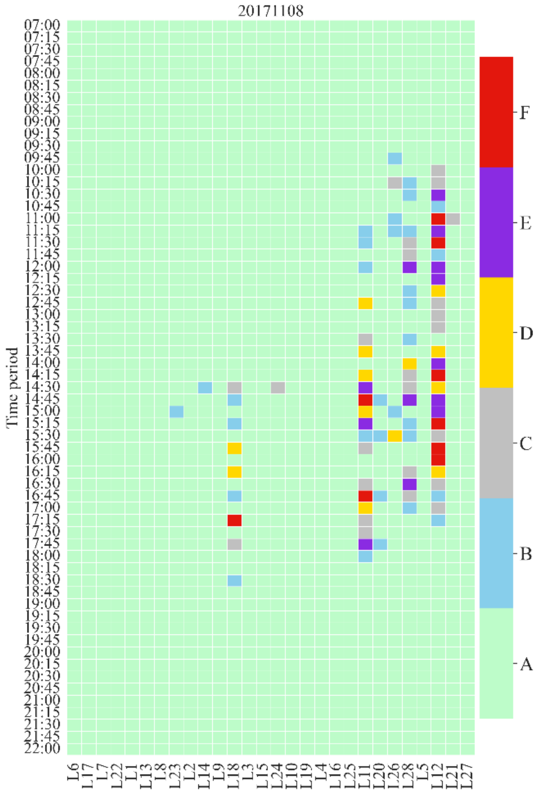

Figure 5.8 PLoS on 2017/11/08 (Wednesday)

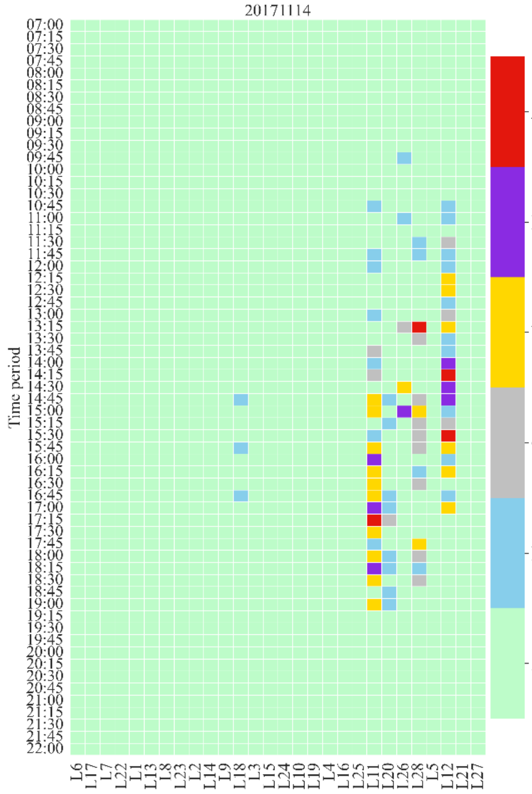

Figure 5.9 PLoS on 2017/11/14 (Tuesday)

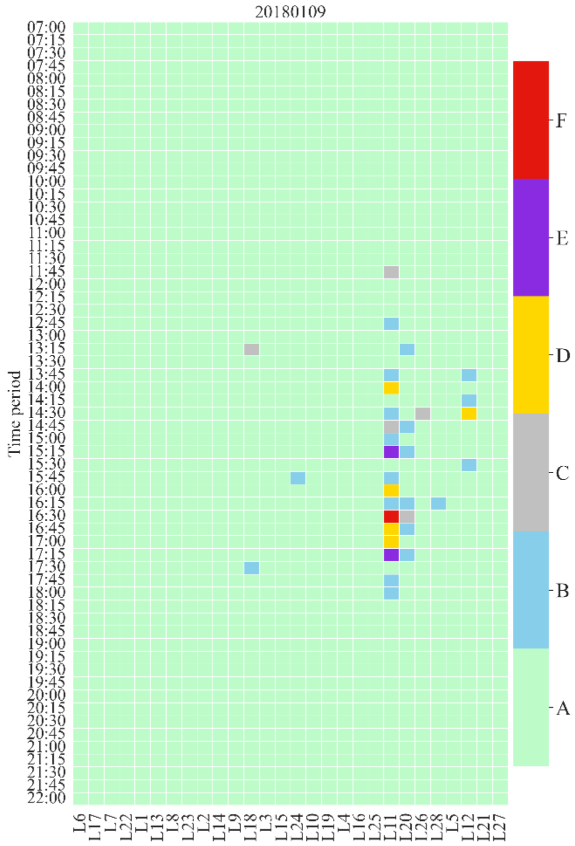

Figure 5.10 PLoS on 2018/01/09 (Tuesday)

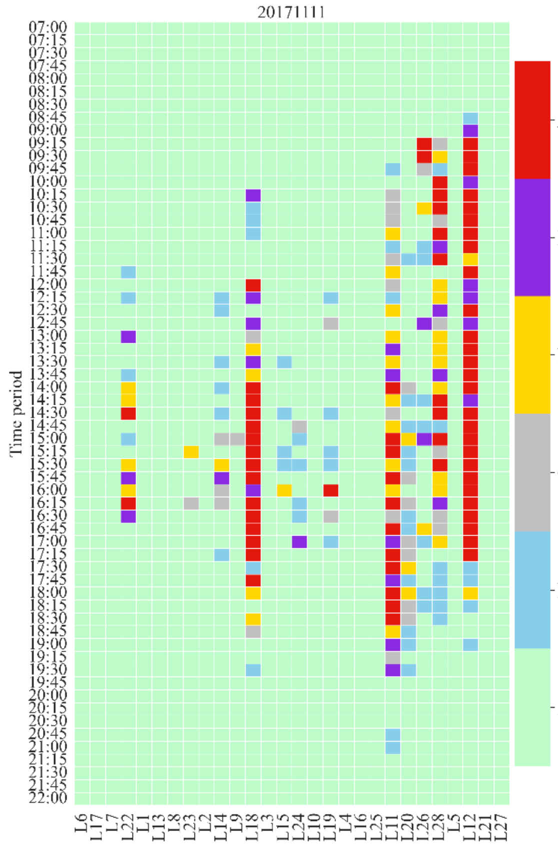

Figure 5.11 PLoS on2017/11/11(Saturday)

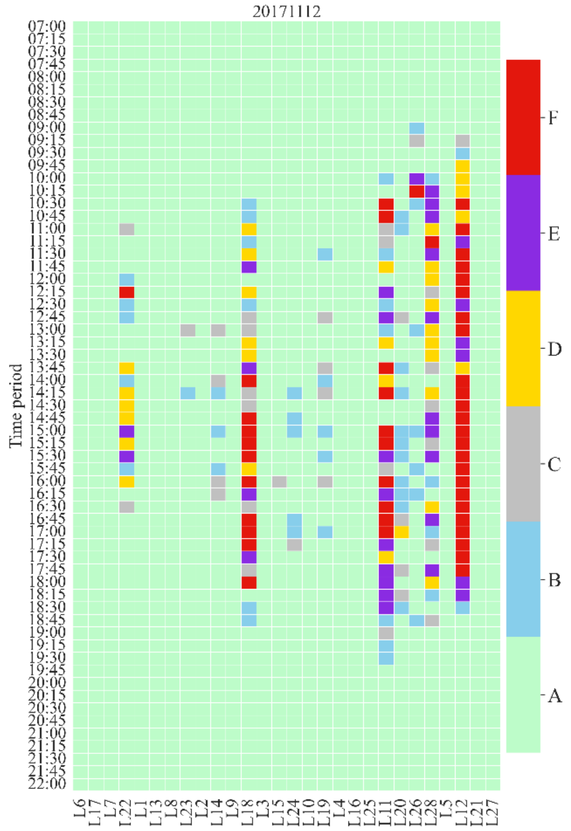

Figure 5.12 PLoS on 2017/11/12(Sunday)

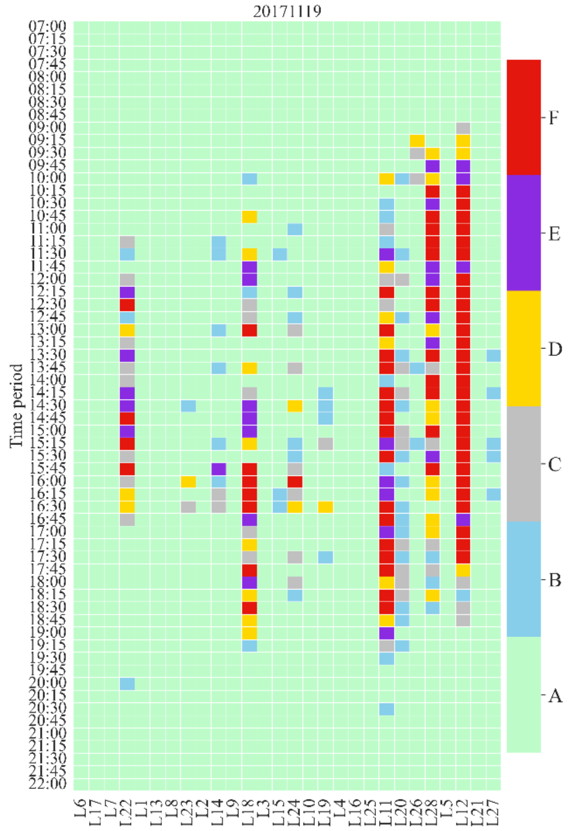

Figure 5.13 PLoS on 2017/11/19 (Sunday)

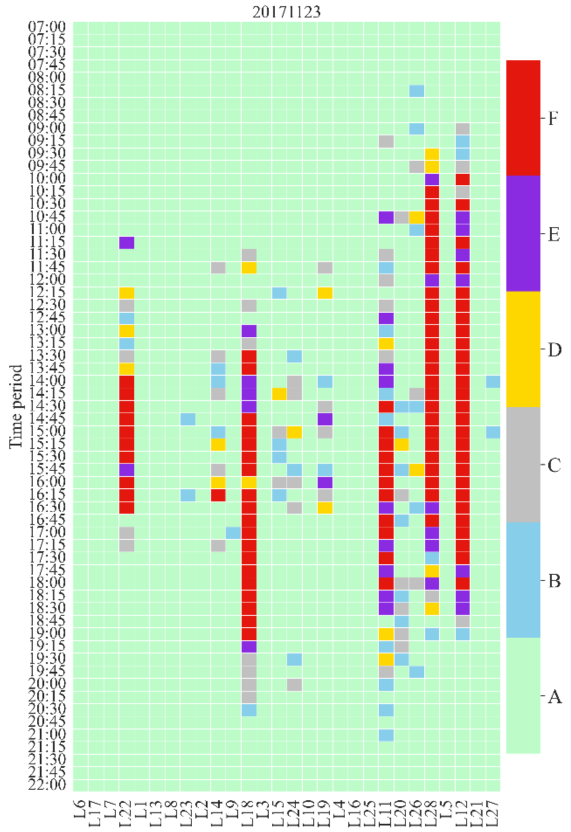

Figure 5.14 PLoS on 2017/11/23 (Thursday, Holiday)

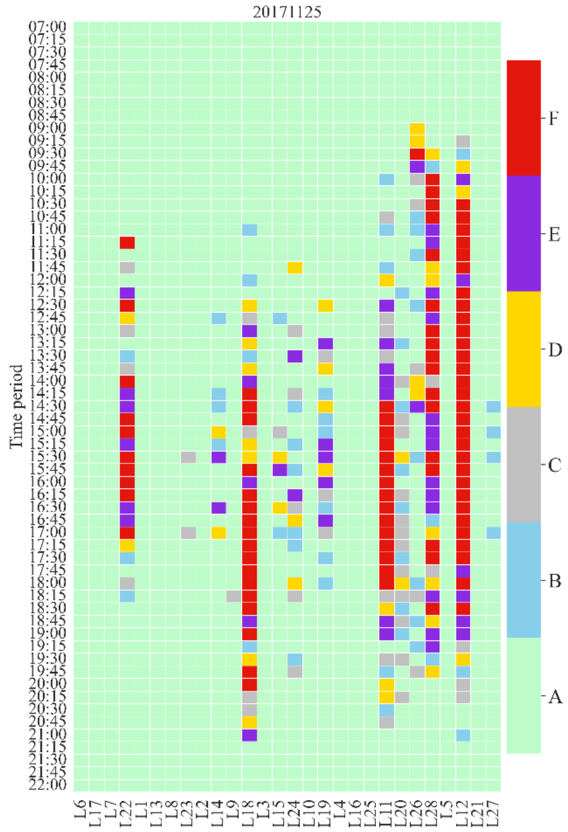

Figure 5.15 PLoS on 2017/11/25 (Saturday)

5.4.2 The PLoS under different scenario

1 Illumination event

To meet with the tourism demand, many of the temples and shrines in Kyoto held illumination events at night. The historic buildings and gardens are illuminated so that visitors can enjoy the very beautiful view of these places extremely with maples in autumn and cherry blossoms in spring. There are three famous temples (Kodai Temple, Entoku Temple and Kiyomizu Temple) that held illumination events in the study area shown by arrows in Figure 5.16. The illumination events were held during 2017/11/11~12/3 from 17:00 to 21:30.

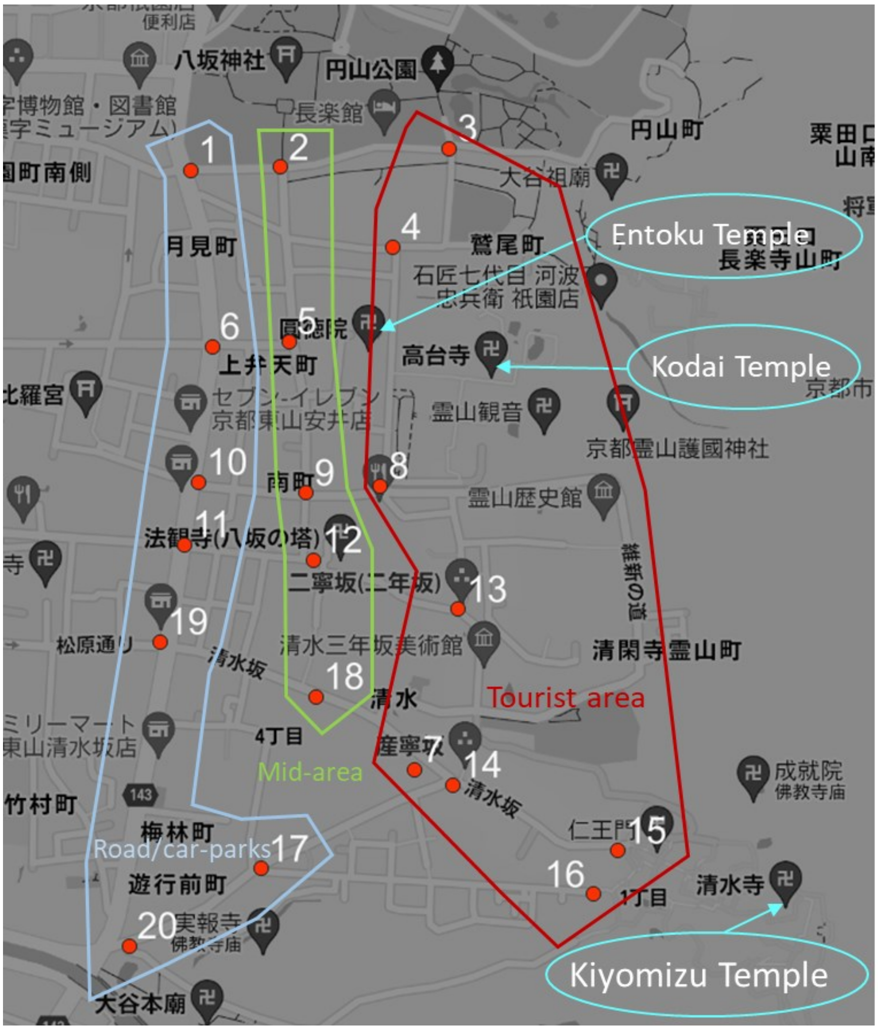

Figure 5.16 Area classification and illumination place

Figures 5.17 and 5.19 represent the PLoS when there had an illumination event and Figures 5.18 and 5.20 represent the PLoS when there had no illumination event. When comparing the PLoS with and without event, it can be recognised that the PLoS of some links (link L18, link

L11, link L28 and link L12) ranked relatively high even after 18:00 (black dotted line) when there had an illumination event while the PLoS was very low after 18:00 if there had no illumination event. There are parking service for private cars and bus stops along the main road between sensor 1 and 20. A major railway station-Kyoto station is in the lower left direction of sensor 20. Links L18 and L28 are near the illumination place and links L11 and L12 are the main corridors to enter and exit the study area. These links are consistent with the frequent patterns identified in chapter 6.

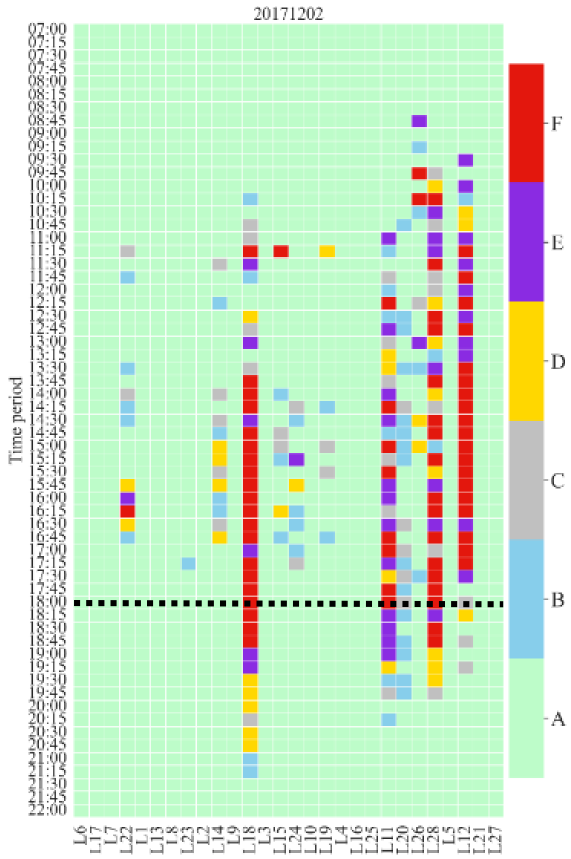

Figure 5.17 PLoS on 2017/12/02 (Saturday)

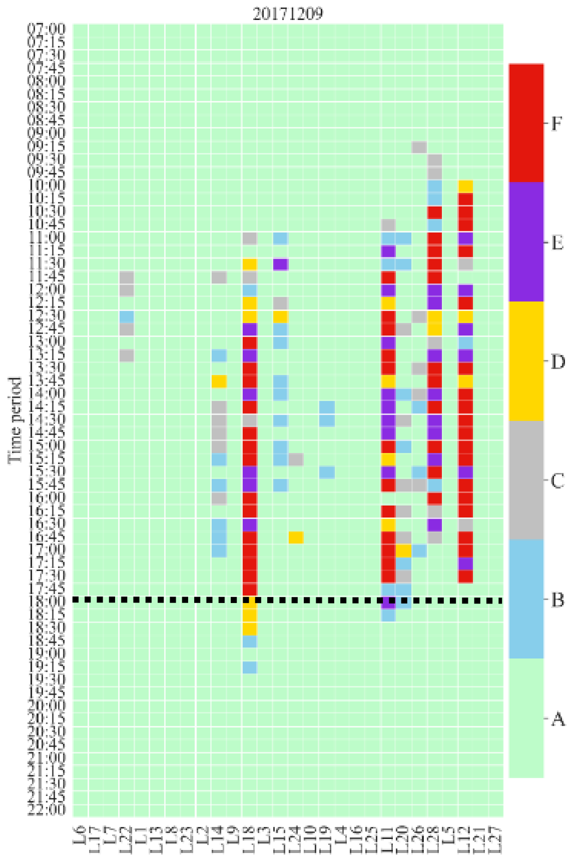

Figure 5.18 PLoS on 2017/12/09 (Saturday)

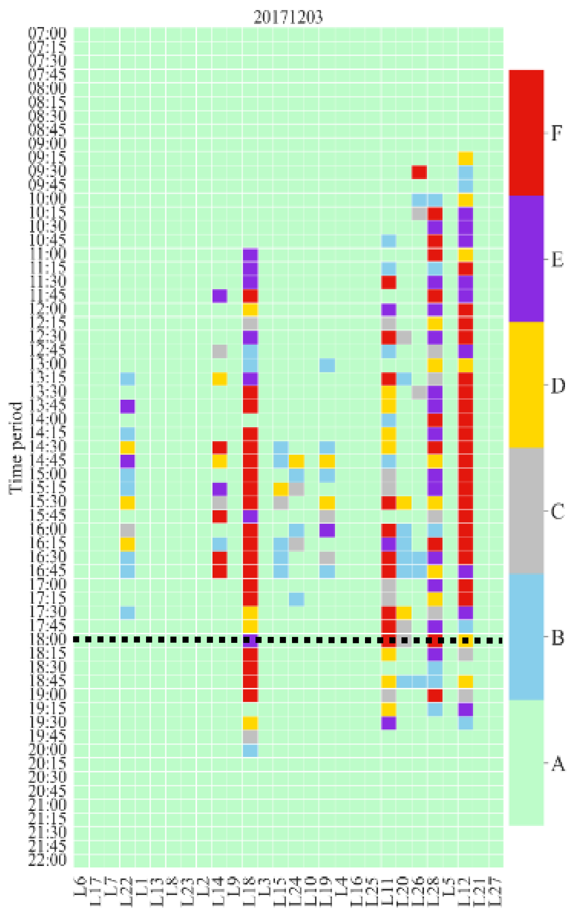

Figure 5.19 PLoS on 2017/12/03 (Sunday)

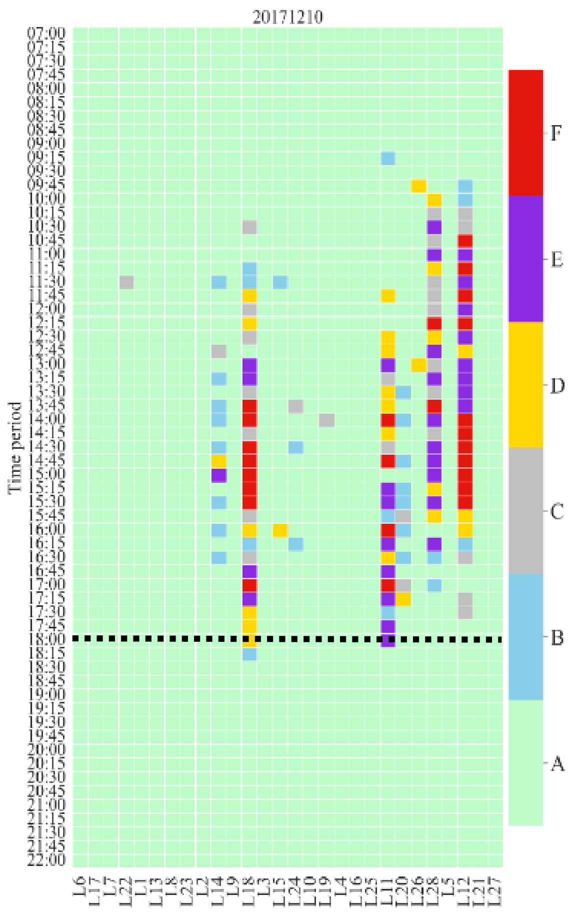

Figure 5.20 PLoS on 2017/12/10 (Sunday)

Based on the characteristics of the locations where the Wi-Fi packet sensor was installed, the whole study area can be classified into three groups named 'Road / Car Parks', 'Mid-area' and 'Tourist area' as shown in Figure 5.16. Figure 5.21 represents the estimated pedestrian flow distribution at different areas over time. The aggregation time period is 1 hour. The estimated pedestrian flow is real pedestrian flow estimated based on the Wi-Fi packet sensor observation and calculated based on the equations (4.1) and (4.2) and parameters in chapter 4.

We can see the observations in 'tourist area' rapidly decreased after 17:00 without the event, and the number of observations in 'Road / Car parks' becomes larger than those in 'Tourist area' from 17:00. However, when there had an illumination event, the decreasing speed of pedestrians in 'Tourist area' is slow and the pedestrian in 'Tourist area' kept larger than that in Road/Car parks area. Therefore, it can be said that the leaving time of the tourists becomes later with the illumination event.

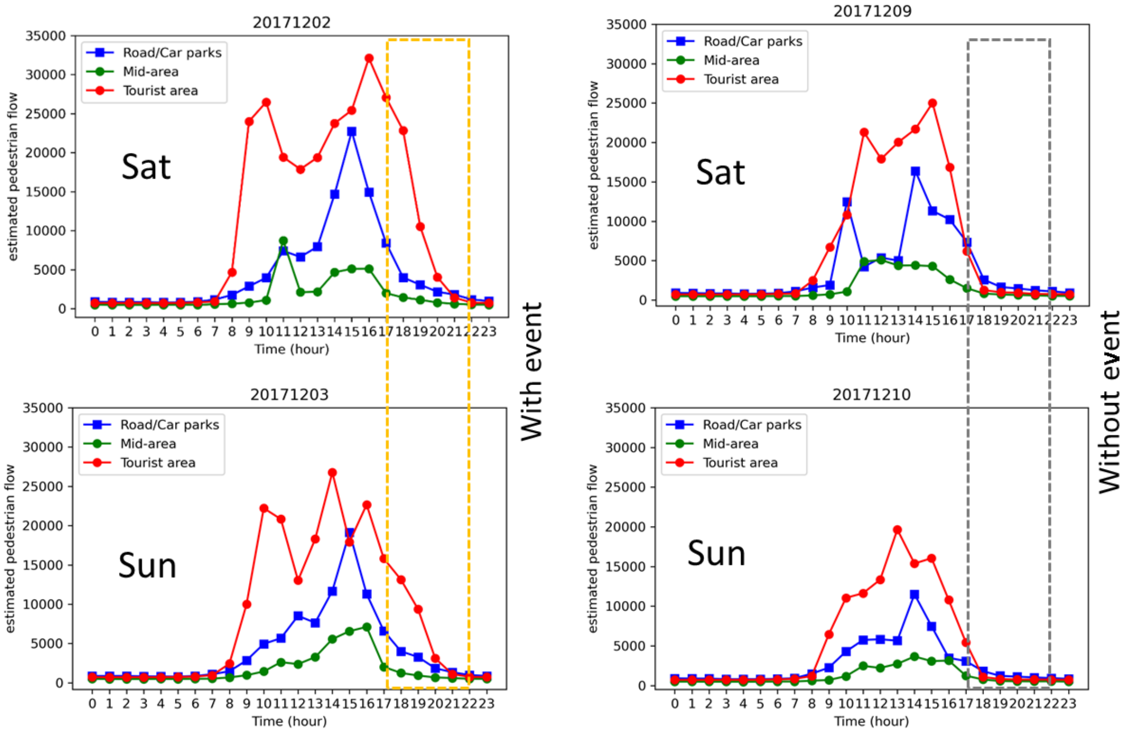

Figure 5.21 Distribution of estimated real pedestrian flow of different areas

2 The PLoS on New Year's Day

Figures 5.22 and 5.23 show the PLoS distribution on Japanese New Year's Eve and New Year's Day. From Figures 5.22 and 5.23 it can be seen that the PLoS of link L22 (near Yasaka shrine) becomes high after 22:00 and link L6(near Yasaka shrine) and link L18 (near Kodai Temple) also had high PLoS level during the start of the New Year. This may be because of the Hatsumode activity (初詣 in Japanese). Hatsumode is a tradition of Japan and it refers to the first visit to a shrine or a temple in the New Year. On this occasion, people pray in the hopes of having a good year ahead. The activity is public and open to anyone. People visit on the first, second or third day of the year as most are on vacation on those days. Generally, most people visit on the first day.

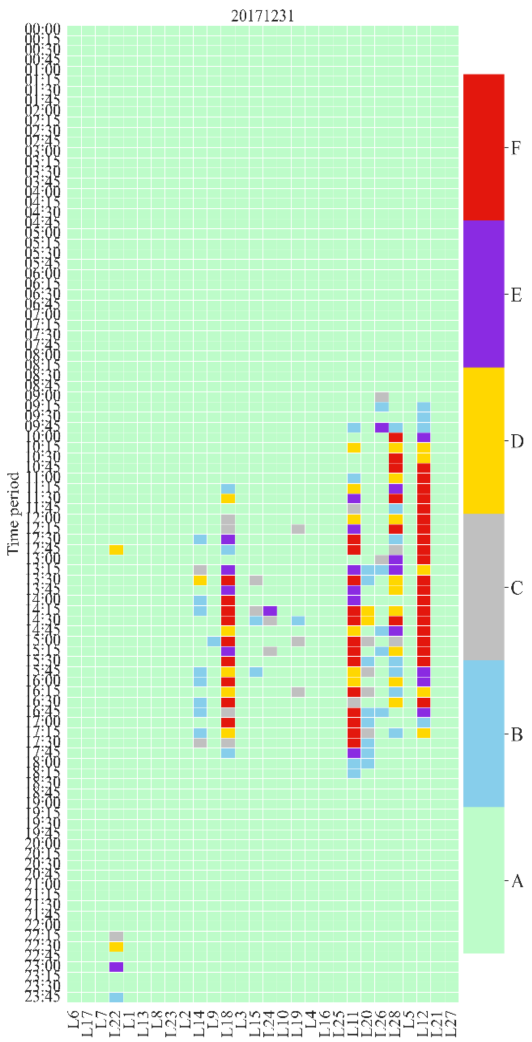

Figure 5.22 PLoS on 2017/12/31 (New Year's Eve)

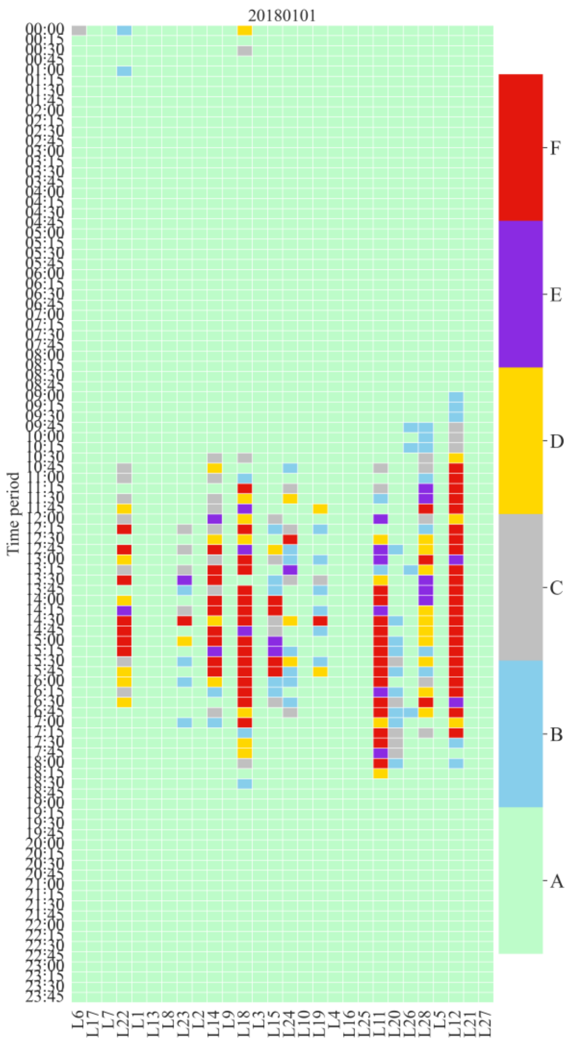

Figure 5.23 PLoS on 2018/01/01 (New Year's Day)

5.5 **Discussion**

Through analysis above, it can be known that the links L12, L28, L11, L18 have rather higher PLoS ranking than other places. These links are consistent with the frequent patterns in chapter 6. Because these places were always at level E or F, meaning that the tourists may always suffer from crowding which may make them uncomfortable. Therefore, the local government or destination management organisation managers should take measures to balance pedestrian flow on these streets so as to improve the service. For example, like Figure 5.24 shows, when links L18 and L11 were busy but the surrounding link like links L19 and L10 were not busy, tourism managers can provide such information to tourists to help them make better planning of sightseeing routes or take measures (such as roundabout or direction rule) to guide them to use other circuitous routes.

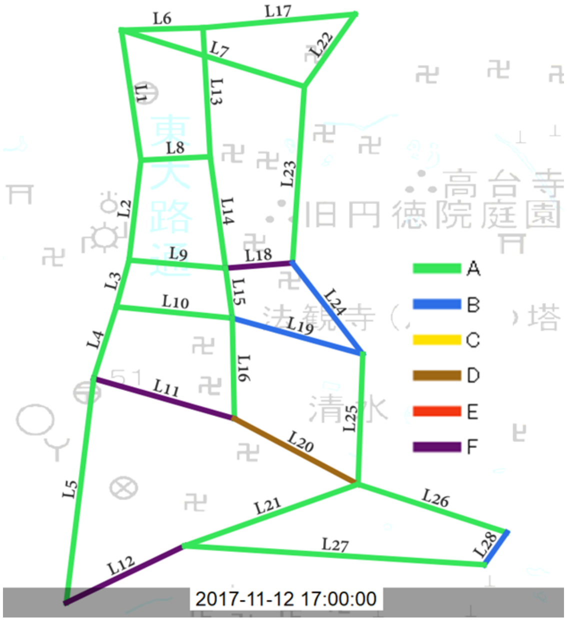

Figure 5.24 PloS level of links

5.6 **Conclusions**

A better understanding of the PLoS in tourist areas can help the managers provide better service and the tourists can have a better experience. In this chapter, the Wi-Fi packet sensor data was used to calculate pedestrian flow rate and evaluate the level of crowding in a tourist area. This data source has an ability to capture the situations of crowd movement in a simple and tractable way. Based on this method, the pedestrian flow can be controlled by time and space. The criteria developed by HCM (2000) was used as a benchmark and pedestrian flow is estimated through matching the same AMACs observed by two sensors located at ends of a link firstly then converting to real pedestrian flow by using the parameters in chapter 4. As a result, it is possible to evaluate the PLoS of a link and provide ideas which street should be improved in specific. It was also recognized that the PLoS of links near illumination event area were ranked higher than without event and people tend to stay longer in tourist area when there had an illumination event, this inspires us holding such an event may help to improve vehicle traffic conditions because the demand will decrease on bus stations or cross-sections near the tourist area. This may be justified by comparing it with traffic detector data in the future.

References

AASHTO, 2021. Guide for the Planning, Design, and Operation of Pedestrian Facilities, 2nd ed. American Association of State Highway and Transportation Officials, Washington, DC.

Aghaabbasi, M., Moeinaddini, M., Shah, M., Asadi-Shekari, Z., 2016. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. Journal of Transport & Health. https://doi.org/10.1016/j.jth.2016.08.012

Ahmed, T., Moeinaddini, M., Almoshaogeh, M., Jamal, A., Nawaz, I., Alharbi, F., 2021. A New Pedestrian Crossing Level of Service (PCLOS) Method for Promoting Safe Pedestrian Crossing in Urban Areas. IJERPH 18, 8813. https://doi.org/10.3390/ijerph18168813

Al-Azzawi, M., 2004. Factors affecting pedestrian walking speeds (PhD Thesis). Napier University.

Al-Mukaram, N., Musa, S.S., 2020. Determination of Pedestrian Level of Service on Sidewalks in Samawah City. IOP Conf. Ser.: Mater. Sci. Eng. 928, 022077. https://doi.org/10.1088/1757-899X/928/2/022077

Anciaes, P.R., Jones, P., 2018. Estimating preferences for different types of pedestrian crossing facilities. Transportation Research Part F: Traffic Psychology and Behaviour 52, 222–237. https://doi.org/10.1016/j.trf.2017.11.025

Asadi-Shekari, Z., Moeinaddini, M., Zaly Shah, M., 2014. A pedestrian level of service method for evaluating and promoting walking facilities on campus streets. Land Use Policy 38, 175–193. https://doi.org/10.1016/j.landusepol.2013.11.007

Asadi-Shekari, Z., Moeinaddini, M., Zaly Shah, M., 2013a. Non-motorised Level of Service: Addressing Challenges in Pedestrian and Bicycle Level of Service. Transport Reviews 33, 166–194. https://doi.org/10.1080/01441647.2013.775613

Asadi-Shekari, Z., Moeinaddini, M., Zaly Shah, M., 2013b. Disabled Pedestrian Level of Service Method for Evaluating and Promoting Inclusive Walking Facilities on Urban Streets. Journal of Transportation Engineering 139, 181–192. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000492

Banerjee, A., Maurya, A.K., Lämmel, G., 2018. A review of Pedestrian flow characteristics and level of service over different pedestrian facilities. Coll Dyn 3, A17. https://doi.org/10.17815/CD.2018.17

Croft, P., Elazar, N., Levasseur, M., 2013. Guide information for pedestrian facilities.

Dixon, L., 1996. Bicycle and Pedestrian Level-of-Service Performance Measures and Standards for Congestion Management Systems. https://doi.org/10.3141/1538-01

Fruin J.J., 1971. Pedestrian planning and design. Metropolitan Association of Urban Designers and Environmental Planners.

Gallin, N., 2001. Quantifying pedestrian friendliness: guidelines for assessing pedestrian level of service. Road and Transport Research 10.

Ghani, N.A., Shimizu, T., Mokhtar, S., 2015. Assessment of Pedestrian Facilities in Malacca World Heritage Site, Malaysia using P-Index Method. Journal of the Eastern Asia Society for Transportation Studies 11, 1535–1554. https://doi.org/10.11175/easts.11.1535

Gr, B., Parida, P., Advani, M., Parida, M., 2018. Pedestrian level of service model for evaluating and improving sidewalks from various land uses. European Transport.

HCM, 2000. Highway Capacity Manual 2000. Transportation Research Board, Washington, D. C.

He, L., Lin, X., Liu, Q., Tao, J.X., 2020. A numerical model for impacts of left-turn non-motorized vehicles on through lane capacity metrics. https://doi.org/10.5604/01.3001.0014.4199

Kadali, B.R., Vedagiri, P., 2015. Evaluation of pedestrian crosswalk level of service (LOS) in perspective of type of land-use. Transportation Research Part A: Policy and Practice 73, 113–124. https://doi.org/10.1016/j.tra.2015.01.009

Khisty, C.J., 1994. Evaluation of pedestrian facilities: beyond the level of service concept. Transportation Research Board 45–50.

Krupat, E., 1985. People in cities: The urban environment and its effects, People in cities: The urban environment and its effects. Cambridge University Press, New York, NY, US.

Landis, B., Vattikuti, V.R., Ottenberg, R.M., McLeod, D.S., Guttenplan, M., 2001. Modeling the Roadside Walking Environment: Pedestrian Level of Service. https://doi.org/10.3141/1773-10

Marisamynathan, S., Vedagiri, P., 2019. Pedestrian perception-based level-of-service model at signalized intersection crosswalks. J. Mod. Transport. 27, 266–281. https://doi.org/10.1007/s40534-019-00196-5

McKinsey & Company, & World Travel & Tourism Council, 2017. Coping with success: Managing overcrowding in tourism destinations.

Molyneaux, N., Scarinci, R., Bierlaire, M., 2021. Design and analysis of control strategies for pedestrian

flows. Transportation 48, 1767–1807. https://doi.org/10.1007/s11116-020-10111-1
Mōri, M., Tsukaguchi, H., 1987a. A new method for evaluation of level of service in pedestrian facilities. Transportation Research Part A: General 21, 223–234. https://doi.org/10.1016/0191-2607(87)90016-1
Muraleetharan, T., Hagiwara, T., Adachi, T., Kagaya, S., 2005. Method to determine pedestrian level-of-service for crosswalks at urban intersections. Journal of the Eastern Asia Society for Transportation Studies 6, 10.
Otak, I., 1997. Pedestrian Facilities Guidebook 248.
Parida, P., Najamuddin, Parida, M., 2007. Development of qualitative evaluation methodology for sidewalks in Delhi. ITPI JOURNAL 27–33.
Raad, N., Burke, M., 2017. Pedestrian Levels-of-Service tools: problems of conception, factor identification, measurement and usefulness, in: 39th Australasian Transport Research Forum (ATRF 2017). Auckland.
Radisya Pratiwi, A., Zhao, S., Mi, X., 2015. Quantifying the relationship between visitor satisfaction and perceived accessibility to pedestrian spaces on festival days. Frontiers of Architectural Research 4, 285–295. https://doi.org/10.1016/j.foar.2015.06.004
Rahul, T.M., Manoj, M., 2020. Categorization of pedestrian level of service perceptions and accounting its response heterogeneity and latent correlation on travel decisions. Transportation Research Part A: Policy and Practice 142, 40–55. https://doi.org/10.1016/j.tra.2020.10.011
Sarkar, S., 1993. Determination of service levels for pedestrains with european examples. Transportation Research Record 35–42.
Shamsuddin, S., Ujang, N., 2008. Making places: The role of attachment in creating the sense of place for traditional streets in Malaysia. Habitat International 32, 399–409. https://doi.org/10.1016/j.habitatint.2008.01.004
Talavera-Garcia, R., Soria-Lara, J., 2015. Q-PLOS, developing an alternative walking index. A method based on urban design quality. Cities 45, 7–17. https://doi.org/10.1016/j.cities.2015.03.003
Tan, D., Wang, W., Lu, J., Bian, Y., 2007. Research on Methods of Assessing Pedestrian Level of Service for Sidewalk. Journal of Transportation Systems Engineering and Information Technology 7, 74–79. https://doi.org/10.1016/S1570-6672(07)60041-5
Teknomo, K., Takeyama, Y., Inamura, H., 2000. Determination of Pedestrian Flow Performance Based on Video Tracking and Microscopic Simulations 23, 5.
Transportation Research Board, 2016. Highway Capacity Manual: A Guide for Multimodal Mobility Analysis, 6th ed. The National Academies Press, Washington, DC.
Ujjwal, J., Bandyopadhyaya, R., 2021a. Development of Pedestrian Level of Service (PLOS) model and satisfaction perception rating models for pedestrian infrastructure for mixed land-use urban areas. Transportation. https://doi.org/10.1007/s11116-021-10247-8
Wen, Y., Yan, K., Yu, C., 2012. Level of Service Standards for Pedestrian Facilities in Shanghai Metro Stations 2072–2078. https://doi.org/10.1061/40932(246)339
Zahid, M., Chen, Y., Jamal, A., Al-Ofi, K.A., Al-Ahmadi, H.M., 2020. Adopting Machine Learning and Spatial Analysis Techniques for Driver Risk Assessment: Insights from a Case Study. International Journal of Environmental Research and Public Health 17, 5193. https://doi.org/10.3390/ijerph17145193
Zhang, L., Prevedouros, P.D., 2003. Signalized Intersection Level of Service Incorporating Safety Risk. Transportation Research Record 1852, 77–86. https://doi.org/10.3141/1852-11
Zhao, L., Bian, Y., Lu, J., Rong, J., 2014. Method to Determine Pedestrian Level of Service for the Overall Unsignalized Midblock Crossings of Road Segments. Advances in Mechanical Engineering 2014, 1–9. https://doi.org/10.1155/2014/652986

Chapter 6: Identifying golden routes in tourist areas based on Wi-Fi packet sensor

6.1 Introduction and research objective

Tourism is one of the pillar industries of Japanese economic development. As one of the most famous tourist attractions of Japan, Kyoto attracted more than 53 million tourists (more than 7 million coming from overseas) based on the statistical data of 2017. According to the Japan National Tourism Organisation, the number of foreign visitors in 2018 exceeded 31 million. Tourists have not only brought economic growth to Kyoto city but also caused many problems, for example, severely crowded conditions inside buses, discomfort to residents, and heavy traffic congestion. While the Japanese Government continues to develop tourism to promote the economy, how to reasonably guide and manage tourists, to provide a comfortable travel experience to visitors and reduce negative impacts on residents, has become an urgent problem. Therefore, it is increasingly important to analyse and understand movement patterns and behaviours of tourists to alleviate congestion in tourist areas and improve services offered to citizens. This chapter will rely on movement data from Wi-Fi packet sensors. By examining the digital footprint of pedestrians, this research is aimed at monitoring crowds' movement behaviour in a small area and network.

One of the most important topics in pedestrian flow research is movement pattern analysis. Selecting the appropriate data mining technique is essential for detecting trajectory patterns in a complex dataset. Sequential pattern mining (SPM) is a topic of data mining used to identify frequent subsequences as patterns in a sequence database. Examples of sequence data include text, DNA sequences, web usage data, and multiplayer games. SPM can be applied to many domains, including discovering customer buying patterns in retail stores, identifying plan failures, and finding network alarms. In this chapter, by analysing sequential trajectory data from tourists in the massive tourist area of Kyoto, Japan, it can identify their frequently used routes and thereby develop strategies to manage crowds and alleviate congestion in the tourist area to make it more convenient and prevent crowd-induced disasters. The extracted knowledge can also be helpful to public space design.

The remainder of this paper is organised as follows. In section 6.2, a brief overview of current research on this topic was provided. Section 6.3 describes the data collection procedure and the research area as well as the framework of this chapter. Section 6.4 aims to explore the movement characteristics of tourists based on Wi-Fi packet sensor data. K-means clustering analysis is applied to extract the tourist-like probe data. Section 6.5 aims to identify a golden route that is most frequently used based on SPM analysis. Finally, section 6.6 concludes the chapter.

6.2 **Related research**

SPM is one of the most popular data mining tasks on sequences. It consists of discovering interesting subsequences in a set of sequences, where the interest of a subsequence can be measured in terms of various criteria, such as its occurrence frequency, length, and profit.

Table 6.1 Applications of sequential pattern mining.

Studies	Application area
Wang et al. (2007)	Bioinformatics
Fournier-Viger et al. (2008); Ziebarth et al. (2015)	e-Learning
Srikant and Agrawal (1996)	Market basket analysis
Pokou et al. (2016)	Text analysis
Schweizer et al. (2015)	Energy reduction in smart homes
Fournier-Viger et al. (2012)	Webpage click-stream analysis
Reps et al. (2012); Batal. et al. (2011); McAullay et al. (2005); Norén et al. (2008); Jin et al. (2008); Wright et al. (2015)	Healthcare and medical fields

The SPM problem was first introduced in 1995 by Agrawal and was defined as follows: 'Given a database of sequences, where each sequence consists of a list of transactions ordered by transaction time and each transaction is a set of items, SPM is used to discover all sequential patterns with user-specified minimum support, where the support of a pattern is the number of data-sequences that contain the pattern' (Agrawal and Srikant, 1995). Its original applications were in the retail industry where it was used to predict whether a customer is likely to purchase its sequel within some time period after purchasing a certain book. SPM has numerous real-life applications since data are naturally encoded as sequences of symbols in many fields, such as those listed in Table 6.1.

Several previous studies in the transport and travel sciences have extracted information from massive trajectory databases using SPM (Bermingham and Lee, 2014). Moreover, to understand travellers' behaviour, pattern mining techniques are suitable for detecting hidden patterns in extensive databases, which are relevant to tourism and destination management research (Mooney and Roddick, 2013). By applying SPM to the tourist trajectory, we can identify relationships between one tourism spot and others to manage or propose planning for tourism development. Bin et al. (2019) summarised the SPM algorithm in route recommendations in tourism studies. Using data collected from websites, they proposed a method to integrate heterogeneous tourism data to construct a point-of-interest (POI) knowledgebase and massive structured POI visit sequences. They also proposed the POI-Visit SPM algorithm to generate fine-grained POI routes. The feasibility and effectiveness of the algorithm were demonstrated using a real-life tourism dataset. These studies indicate the feasibility of applying SPM to tourism planning and management.

In this chapter, I try to identify the frequent trip patterns of tourists by an SPM method using data collected by the emerging Wi-Fi sensing technology.

6.3 **Methodologies**

6.3.1 Research Area

The data used in this chapter were collected from the Higashiyama area around Kiyomizu Temple, which is one of the busiest tourist areas in Kyoto city, Japan. The detection area was about 0.6 km^2 (1,000-metre long and 600-metre wide). Figure 6.1 shows the study area on the map. 20 sensors were equipped to collect data for 6 months (from 2017/10/1 to 2018/3/28). Table 6.2 lists the sensor IDs, and the observation location names.

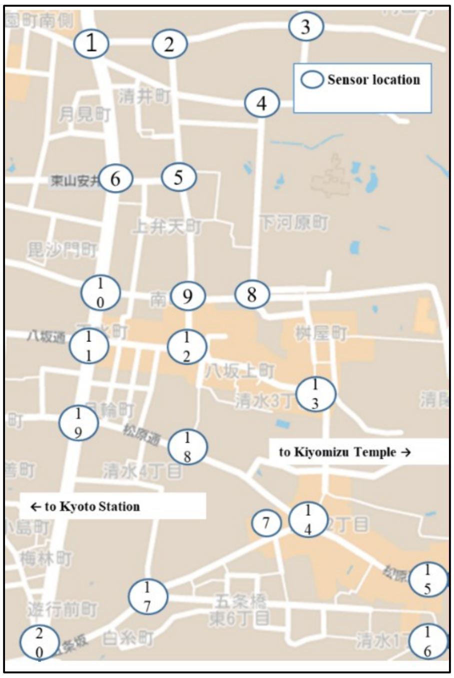

Figure 6.1 Coverage of observation locations in Higashiyama.

Table 6.2 Observation locations.

Sensor ID	Location	Sensor ID	Location
1	Higashioji Shinkomichi	11	Higashioji Yasakatsu
2	Shimokawara Shinkomichi	12	Shimokawara Yasakatsu
3	Nenenomichi Shinkomichi	13	Nineizaka Yasakatsu
4	Nenenomichi Chuo	14	Sanneizaka Matsubaradori
5	Shimogawara Yasui	15	Matsubaradori Kiyomizu Temple
6	Higashioji Yasui	16	Chawanzaka Kiyomizu Temple
7	Shimizuzaka Parking	17	Gojozaka Chawanzaka
8	Nenenomichi Kodaiji	18	Matsubaradori Chuo
9	Shimogawara Kodaiji	19	Higashioji Matsubaradori
10	Higashioji Kodaiji	20	Higashioji Gojozaka Naka

6.3.2 Sequential Pattern Mining Framework

The framework for discovering frequent travel patterns of tourists from Wi-Fi packet collector data is outlined in Figure 6.2. The data cleaning module is used first to remove the randomised AMACs and non-movement data to save calculation time. The clustering analysis module identifies the different smart device users. The travel routes generating module discovers a series of sequential patterns of destinations visited by tourists by adopting the constrained sequential pattern discovery using equivalence classes (CSPADE) algorithm. Finally, the frequently used routes are selected based on support, confidence, and lift. The details of these processes are presented in Sections 6.4 and 6.5.

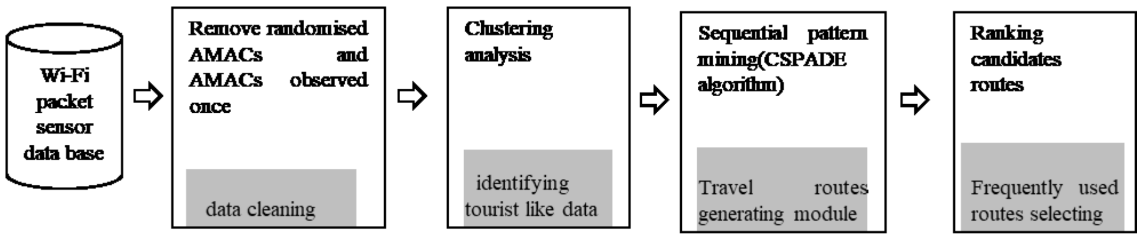

Figure 6.2 Sequential pattern mining framework.

6.4 **Categorisation of observations by clustering analyses**

In this section, I investigate whether it can identify different types of users based on the observed data. Since Wi-Fi packet sensor data cannot identify the characteristics of device users, non-hierarchical clustering was performed to understand different characteristics of observations. Izakian et al. (2016) pointed out that clustering is one of the most powerful techniques to reveal hidden patterns and structures in data. Among many clustering algorithms, unsupervised machine learning algorithms, such as the K-means algorithm, help to avoid the transposition of expectations on clustering results (Jain, 2010). As one of the most popular clustering algorithms, K-means clustering is used for partitioning a given dataset into a set of k groups (i.e., k clusters), where k represents the number of groups. It classifies objects in multiple groups, such that objects within the same cluster are very similar, whereas objects from different clusters are very different. In other words, it tries to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as Euclidean-based distance (Syakur et al., 2018). K-means clustering analysis was applied here to cluster AMAC addresses, to identify the type of device users. K-means clustering is relatively computationally quick. The approach was also used by Bayarma et al. (2007) to identify groups of travel patterns from travel diary data.

From pedestrian observation data collected from 2017/11/06 (Monday) to 2017/11/12 (Sunday), I removed randomised AMACs and AMACs that were only observed once. A randomised AMAC could be identified through an organisationally unique identifier (OUI). This week was chosen as it is one of the best times to visit Kyoto city because of beautiful weather and Japanese maples. After preprocessing, the data used for clustering analysis are summarised in Table 6.3.

Table 6.3 Data used for clustering analyses.

Total number of records	Number of records after removing random AMACs	Number of records obtained from AMACs observed once	Number of records for analyses	Number of AMACs for analyses
7,642,747	2,292,031	52,014	2,240,017	158,051

The data was aggregated by each AMAC based on the observation time, place, and other indicators. In total, nine factors comprising 57 items were considered for K-means clustering analyses. These factors are described in Table 6.4.

Table 6.4 Factors considered for K-means clustering analyses.

Factors	Definition
Share of observations by each sensor	For an AMAC, the number of observations by a sensor divided by the total observations of this AMAC; one item per sensor installation point.
Share of observations by time	Number of observations each hour divided by the total number of observations; 24 items.
Share of observations by day of week	Number of observations each day of a week divided by the total number of observations; seven items, one for each day of the week.
Observed days	Number of days an AMAC is observed within a week.
Observed sensors	Number of sensors observing an AMAC.
Observed time	Time difference between when an AMAC is first and last observed.
Total observations	Total number of observations for each AMAC.
First observed time	Time when an AMAC is first observed (in hours, dummy variable); 24 items.
Final observed time	Time when AMAC is last observed (in hour, dummy variable); 24 items.

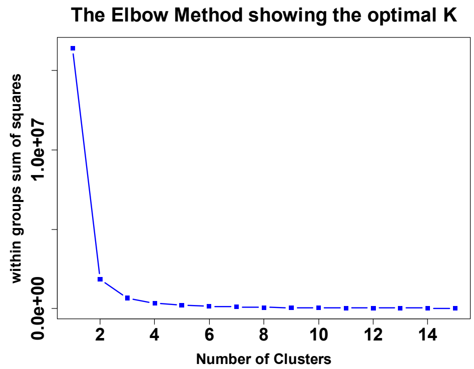

Figure 6.3 Results from the elbow method.

Table 6.5 AMAC count of each cluster.

Cluster	AMAC counts
1	116,448
2	27,962
3	13,536
4	105

To use the K-means clustering method, the number of clusters, K, must be identified in advance. I applied the elbow method to determine the optimal K value. The results of the elbow method are shown in Figure 6.3. The vertical axis corresponds to the sum of the squares of the distances between samples and the centre of their cluster, and the horizontal axis represents the number of clusters. We selected the K value of 4 as the optimal value and used this value for our K-means clustering. The AMAC counts for each cluster are detailed in Table 6.5. Cluster 1 comprised more than 70% of all AMACs.

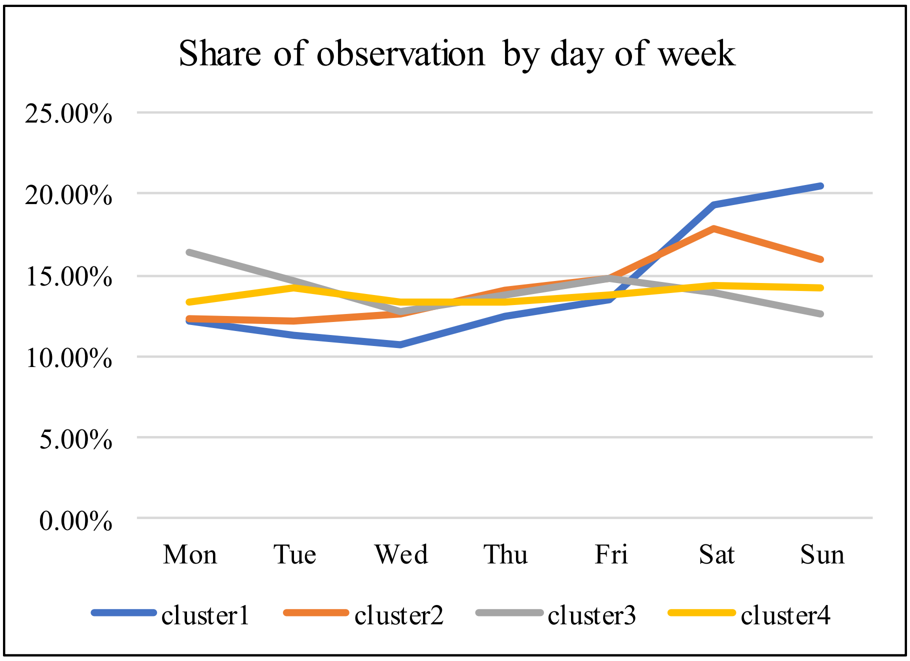

Figure 6.4 Share of observations by the day of the week.

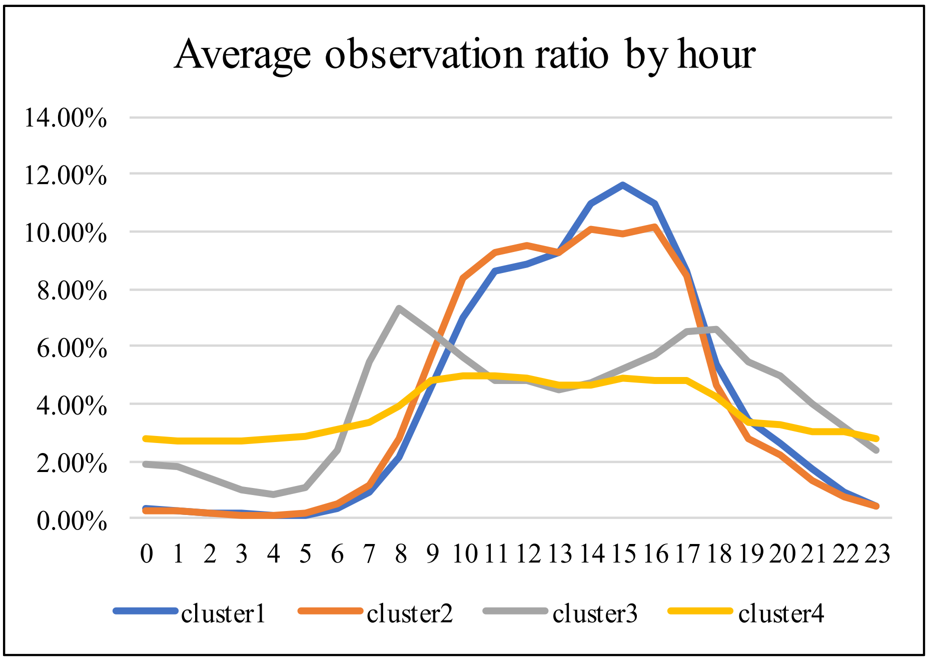

Figure 6.5 Share of observations by the hour.

Figure 6.4 shows the average observation rate with respect to the day of the week, and Figure 6.5 shows the average observation rate with respect to the hour. We can see that cluster 1 had a clear upward trend during weekends and had a peak time between 13:00 and 17:00. Cluster 2 also had a higher observation ratio on weekends than weekdays, but a decreasing trend appeared on Sunday. Cluster 2 had a rather high observation count from 10:00 to 16:00. The average observation ratio of cluster 3 on weekdays was higher than that on weekends and had two peaks at about 8:00 and 18:00. By contrast, there was no obvious change in cluster 4 both weekly and hourly.

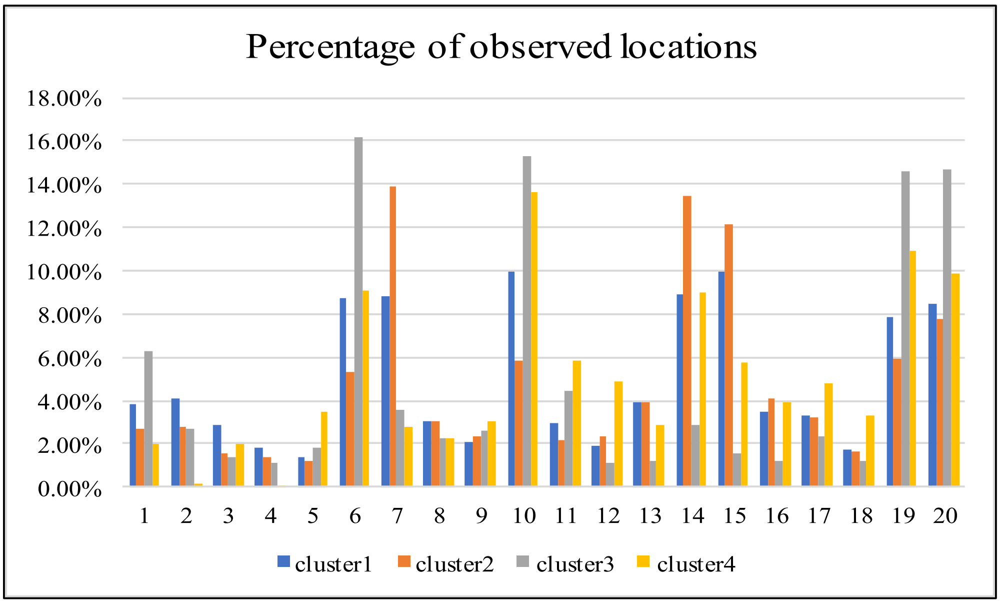

Figure 6.6 Percentage of observed locations

Figure 6.6 shows the percentage of observations in each location (sensor) per cluster. The horizontal axis corresponds to the sensor location (see Table 6.2). Devices in cluster 1 often appeared at sensor locations 6, 7, 10, 14, 15, 19, and 20. Location 7 is a parking place exclusively for sightseeing buses. Moreover, location 14 is a typical access route to Kiyomizu Temple. It is therefore possible to consider that many tourists disembark the sightseeing buses at Shimizuzaka parking and then walk to Kiyomizu Temple via locations 14 and 15. Kiyomizu Temple is located to the right of sensor 15 shown in Figure 6.1. For cluster 2, the sensors at locations 7, 14, and 15 had more observations than other places. Cluster 3 often appeared at locations 6, 10, 19, and 20. Cluster 4 often appeared at locations 10, 19, and 20. Clusters 3 and 4 were often observed at sensors along Higashiojidori (a main road where sensors 1, 6, 10, 11, 19, and 20 were installed). Supermarkets, convenience stores, and apartments are distributed along this street. Furthermore, the Higashiyama Ward office is located near the Higashioji Matsubaradori (sensor 19). These clusters may capture movements, such as daily shopping and business behaviour.

Cluster 1 is considered to include many tourists, especially same-day visitors, because the observation time was several hours and often occurred at weekends, and there were many observations near Kiyomizu Temple, the famous sightseeing spot (a World Heritage Site). As for cluster 2, there were also many observations near Kiyomizu Temple, but the observation time was about 2 days at weekends. A decreasing trend appeared on Sundays, which suggests that some visitors left the detection area on Sundays. Cluster 3 is considered to include travellers who show daily activities (i.e., commuters) because the peak observations occurred in the morning and evening, and there were more observations on weekdays than weekends. From the perspective of spatial distribution, many observations were captured along Higashiojidori, which is one of the main roads connecting the north and south parts of Kyoto city. Cluster 4 is considered to include residents because peak observations occurred more than 5 days per week on average and more than 22 hours per day. Moreover, there were no large fluctuations in the trend of the share of observations according to the day of the week (Figure 6.4) and by time within a day (Figure 6.5). The average number of observed sensors reflects the areas of activity in each different cluster. We can see that users in clusters 1–3 often moved between more than three places, especially in cluster 2, in which users moved to more than seven places. Users in cluster 4 moved to less than three places. Moreover, this cluster had a much higher number of observations for each AMAC. I identified the cluster characteristics based on these analyses (Table 6.6).

Table 6.6 Summary of classification results and estimated attributes.

Classification factor	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Average number of observation days	1.04	1.40	3.63	5.76
Average number of observation hours	2.29	14.41	109.6	158.75
Average number of observation sensors	3.74	7.73	4.89	2.68
Average number of observations	7.02	29.22	30.99	1768.9
Share of observations by time	Peak at 15:00	No obvious peak, distributed uniformly between 10:00 and 16:00	Peak at 8:00 and 18:00	Distributed uniformly all day long
Share of observations by day of week	Weekend > weekday	Weekend > weekday	Weekday > weekend	Little difference
Share of observations by each sensor	Many observations near Kiyomizu Temple and Higashiojidori	Many observations near Kiyomizu Temple	Many observations near Higashiojidori	Many observations near Higashiojidori
Estimated attribute	Same-day visitors	Overnight visitors	Commuters	Residents

6.5 **Extracting frequent trip patterns**

To gain more insight into the movement behaviour of tourists in the Higashiyama area, I applied the SPM technique to AMACs of clusters 1 and 2. In our dataset, a set of pedestrian trips (Wi-Fi probe request observation records) during one day is considered a time series sequence of their visiting destinations. Each time series sequence was assigned a letter that represents the items in that sequence; in our study, our alphabet contained 20 elements; each element was the name of the Wi-Fi packet sensor.

Identifying all frequent sequential patterns in a transaction database, especially in large databases such as those found in healthcare, requires an efficient algorithm to deal with the large search space. Several algorithms, such as SPADE (**S**equential **PA**ttern **D**iscovery using **E**quivalence classes), have been developed. Zaki (2001) proposed this algorithm that uses many strategies to make SPM more efficient. One of the most important features in SPADE is the database layout, which is transformed from horizontal into vertical id-list database format in this algorithm. Another important principle is the equivalent classes. With them, the search space is decomposed into small pieces. Then, these pieces can be processed in the memory independently. Within each equivalence class, depth-first search is used for enumerating the frequent sequences. CSPADE is an extension of the SPADE algorithm and incorporates constraints on sequences, such as lengths or width limitations on the sequences, minimum or maximum gap constraints on consecutive sequence elements, or time windows (Zaki, 2000). CSPADE has been applied to protein folding (Exarchos et al., 2006), hepatitis classification

(Aseervatham and Osmani, 2005), and satellite image processing (Julea et al., 2008). Recently, Ibrahim and Shafiq (2019) used SPM to discover taxi movement patterns in the city of Porto, Portugal. Their dataset included 442 taxi trajectories acquired during one year (Dua and Taniskidou, 2017). I processed the data using the CSPADE algorithm and R package 'arulesSequences' (Buchta et al., 2020).

6.5.1 Definitions and data preparation

In this section, I first introduce some preliminary concepts, and then formulate the SPM problem. As described by Zaki (2000), the problem of mining sequential patterns can be stated as follows: Let $\mathrm{I}=\{\mathrm{i}_1, \mathrm{i}_2, ..., \mathrm{i}_m\}$ be a finite set of items, for example, (A, B, C, D, E, F, and G). A subset of I is called an itemset. We also call an event an itemset. An event is a non-empty unordered collection of items. Without loss of generality, we assume the items in an event are sorted in alphabetic order. A sequence $\mathrm{S} = \langle e_{1,} e_{2, ...,} e_n \rangle (e_j \subseteq \mathrm{I})$ is a temporally ordered list of events where each event $e_j\,(1 \le j \le n)$ is an itemset; for example, <(A, B, C), (D, E), (F, G)>. Let S_2 be another sequence denoted as <(A), (E), (G)>. Sequence S_2 is called a subsequence of sequence S since (A) ⊆ (A, B, C) and (E) ⊆ (D, E) and (G) ⊆ (F, G). The length of the sequence is the total number of items in the sequence. For example, the length of sequence $\mathrm{S} = <(A,\ B,\ C),\ (D,\ E),\ (F,\ G)>$ is 7, and the length of $\mathrm{S}_2\ =\ <(A),\ (E),\ (G)>$ is 3. To facilitate understanding, imagine that customer 1 in Table 6.7 bought items A and B at time 10, then bought item B at time 20, and finally bought items A and B at time 30. In addition, customer 2 bought items A and C at time 20, then bought items A–C at time 30, and finally bought item B at time 50. Comparing their sequences, the subsequence AB → B seems to be frequent.

Table 6.7 Example sequence of shopping database.

SID	EID	Items
Customer 1	10	AB
	20	B
	30	AB
Customer 2	20	AC
	30	ABC
	50	B

Table 6.8 Example sequence of Wi-Fi database.

SID	EID	Location
AMAC 1	10	S1
	30	S3
	40	S4
AMAC 2	30	S1
	40	S2
	50	S4

Similarly, as shown in Table 6.8, we can recognise that AMAC 1 was observed by sensor S1 at time 10, then observed by sensor S3 at time 30, and finally observed by sensor S4 at time

40. AMAC 2 was observed by sensor S1 at time 30, then observed by sensor S2 at time 40, and finally observed by sensor S4 at time 50. The moving sequence from S1 to S4 is frequent for both AMACs in this example. Actually, some AMACs were observed by two different sensors at the same time. The reason may be because of the overlap of the detection areas of the two sensors. To cope with this situation, I filtered the AMAC record by RSSI. A record having a weaker RSSI means it was farther from the sensor, and thus I removed it.

Table 6.9 Definitions and calculation method.

Support	$Support(X,Y) = \frac{number\ of\ patterns\ satisfying\ X\ and\ Y}{number\ of\ total\ patterns}$
Confidence	$Confidence(X,Y) = \frac{number\ of\ patterns\ satisfying\ X\ and\ Y}{number\ of\ patterns\ satisfying\ X}$
Lift	$Lift(X,Y) = \frac{Confidence(X,Y)}{Support(Y)}$

As described before, Wi-Fi packet sensor data include AMAC, timestamp, and sensor ID. It is a kind of sequential data in which each record is represented as a sequence of 'events'. The transaction is the minimum processing unit; in my dataset, one packet record is treated as a transaction. The rules for expressing frequent patterns are called association rules, and the extracted patterns are evaluated by the association rules. Support, confidence, and lift are given as indicators of the association rule. The calculation methods of the indicators are presented in Table 6.9, where X corresponds to an intermediate sequence and Y to a final destination. The task of SPM is to identify frequent sequences, where frequency is defined as having support above a user-defined threshold.

Support indicates the rate of occurrence of a specific trip pattern (X, Y) out of all possible trip patterns. In other words, it is the occurrence probability of the trip pattern (X, Y). *Confidence* indicates the probability of going to destination Y when X is observed. The closer the confidence to 1, the more likely Y will occur after X occurs. *Lift* indicates the correlation between X and Y. If the lift equals 1, it means that X and Y are independent. If the lift exceeds 1, it means that X and Y are positively correlated. For example, consider the rule ($X \Rightarrow Y$) with a lift value greater than 1, i.e., the occurrence of X is positively correlated with the occurrence of Y. If X increases, then the occurrence of Y also increases. If the value of the lift is less than 1, then X and Y are negatively correlated. We can say that a sequence is frequent if its support values are larger than the predetermined threshold. With these definitions and a sufficiently high support threshold, we can identify patterns that more pedestrians will meet; otherwise, if we set support with too low a threshold, the algorithm could generate many insignificant patterns, which means a lower number of pedestrians meeting a pattern. Our target research area, the Higashiyama area, contains many tourist spots within a small area. There are multiple routes heading to the same destination, and a huge number of movement patterns may occur because of sightseeing excursions. Therefore, the threshold for support was set as 0.001, which means 1 out of 1000 people will meet that pattern. The movement pattern is defined as a typical pattern if its support exceeds 0.001.

The gap value is an important indicator of travel patterns. Notably, Wi-Fi packet sensor data, which is a kind of location-dependent data, cannot observe pedestrian data if there is no sensor at a location. This is an issue, for example, when a movement pattern from sensor X to Y is confirmed because it is impossible to distinguish whether the movement is from sensor X to sensor Y directly, or there is a stop in-between at location Z where there is no sensor. Staying behaviour is more likely to happen as the moving time between two points increases. Therefore, in this chapter, considering some tourists may stop to take a photograph or visit other places between locations X and Y, I set the gap time as 1 hour. This means it count the items only when the time difference between transactions is less than 1 hour. Considering the gap time, the case only includes one visit to location Y after location X within 1 hour.

6.5.2 Results of the generated patterns

The data information for the CSPADE algorithm is given in Table 6.10. The results of the algorithm are presented in Table 6.11.

Table 6.10 CSPADE algorithm data.

	AMAC count	Packets count	Average chain length
Cluster 1	116,448	817,720	7.02
Cluster 2	27,962	817,039	29.2

Table 6.11 CSPADE algorithm results.

	Number of patterns	Average support	Average confidence	Average lift
Cluster 1	35,848	0.0028	0.44	1.76
Cluster 2	7,575,202	0.0026	0.49	1.18

The number of patterns extracted for cluster 2 exceeds the AMAC count. This is because multiple patterns were extracted from one AMAC. For cluster 1, the number of patterns is less than the AMAC count. This may be due to many of the tourists moving as a group; cluster 2 has a greater variety of movement patterns than cluster 1.

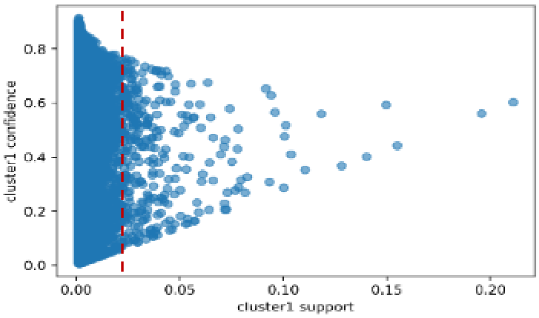

Figure 6.7 Relationship between support and confidence (cluster 1).

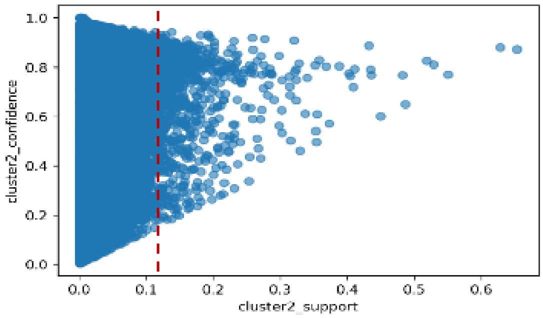

Figure 6.8 Relationship between support and confidence (cluster 2).

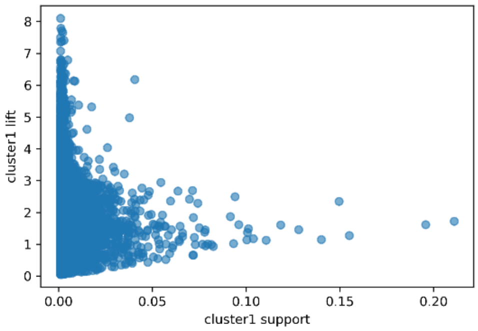

Figure 6.9 Relationship between support and lift (cluster 1).

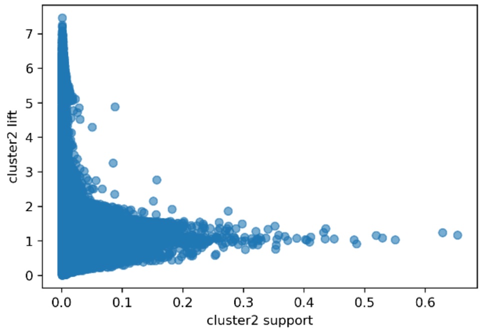

Figure 6.10 Relationship between support and lift (cluster 2).

Figures 6.7 and 6.8 show the relationship between support and confidence of the patterns. The confidence is widely distributed, ranging from about 0 to 1 for each of clusters 1 and 2. Confidence close to 1 indicates that movement patterns always head to a specific location. Figures 6.9 and 6.10 show the relationship between support and lift of the patterns. The lift value is widely distributed, ranging from about 0 to 8 for each of clusters 1 and 2. In addition, the lift value converges to about 1 in the range of high support. Most of the patterns with high confidence and lift value have low values of support. Although it seems to be a typical movement pattern, it cannot be confirmed that it is a golden route, i.e., one that is used most often. By contrast, there are some patterns with moderate support values (after which the scatterplot starts to become sparse) but still relatively high confidence (greater than 0.5) and lift (greater than 1) values.

6.5.3 Extracting golden routes

If pickup patterns are based on only one indicator, most of the movement patterns with high support are between two or three locations, it is hard to say that they represent a typical trip pattern. In addition, a movement pattern with high confidence or lift value has rather low support, so it is uncertain whether many tourists are following the pattern. Therefore, I extracted those patterns whose support, confidence, and lift values exceeded their average values. A total of 3155 patterns (8% of the total) were extracted for cluster 1 and 686,595 patterns (9% of the total) were extracted for cluster 2.

Too many patterns are difficult to display on the map and to confirm that they are typical patterns. By contrast, too few patterns indicate that the extracted patterns do not fully reflect tourist behaviour. Sorting these patterns in descending order of support revealed that many patterns only contain one or two locations, and also many duplicate patterns. To shortlist the generated patterns, I selected the longest trip pattern in the case when there were subsequences.

Table 6.12 Patterns for deriving the golden route of cluster 1.

Patterns	Support	Confidence	Lift
<{"S15"},{"S14"},{"S7"}> => <{"S7"}>	0.0434	0.60	2.39
<{"S7"},{"S14"},{"S15"}> => <{"S15"}>	0.0408	0.72	2.05
<{"S14"},{"S17"}> => <{"S20"}>	0.0389	0.72	2.30
<{"S17"},{"S16"}> => <{"S15"}>	0.0365	0.72	2.07
<{"S20"},{"S17"}> => <{"S16"}>	0.0350	0.54	3.21
<{"S14"},{"S15"},{"S14"}> => <{"S7"}>	0.0345	0.52	2.06
<{"S15"},{"S17"}> => <{"S20"}>	0.0331	0.74	2.35

I only chose patterns if they contained three or more different locations with lift values larger than 1 from sorted patterns. Table 6.12 indicates the seven patterns that were selected for cluster 1. For cluster 2, I stopped at the last pattern in Table 6.13 because there were many duplicate patterns below this pattern (the pattern count was 1127). We define the golden routes as the routes deriving from these patterns (Tables 6.12 and 6.13).

Table 6.13 Patterns for deriving the golden route of cluster 2.

Patterns	Support	Confidence	Lift
<{"S7"},{"S18"}> => <{"S19"}>	0.1036	0.77	1.59
<{"S7"},{"S14"},{"S15"},{"S14"},{"S7"},{"S7"},{"S7"}> => <{"S20"}>	0.0990	0.79	1.36
<{"S20"},{"S17"},{"S16"},{"S16"},{"S15"}> => <{"S15"}>	0.0877	0.88	1.23
<{"S20"},{"S15"},{"S15"},{"S14"},{"S7"}> => <{"S20"}>	0.0875	0.71	1.22
<{"S19"},{"S18"},{"S14"},{"S15"}> => <{"S15"}>	0.0872	0.91	1.27
<{"S15"},{"S14"},{"S14"},{"S7"},{"S7"},{"S7"},{"S17"}> => <{"S20"}>	0.0869	0.91	1.57
<{"S15"},{"S14"},{"S14"},{"S13"}> => <{"S8"}>	0.0850	0.61	1.97

Based on Buchta et al. (2020), the CSPADE rules result is shown as < >, where each item of the pattern is displayed inside {" "}, the comma between {" "} separates items that appear one after another, and the arrow (=>) points to the destination of the pattern. Figures 6.11 and 6.12 help to visualise the generated golden route.

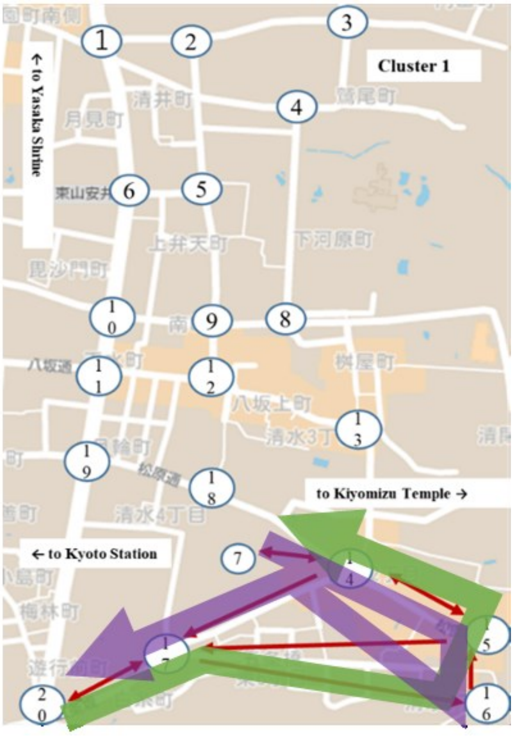

Figure 6.11 Golden route of cluster 1.

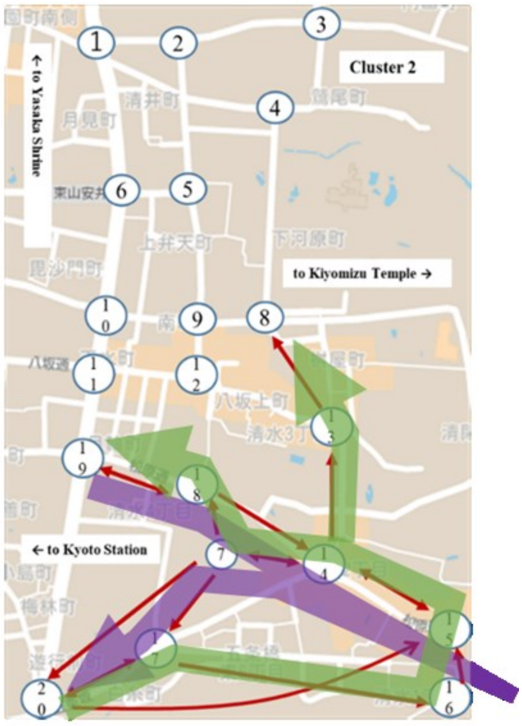

Figure 6.12 Golden route of cluster 2.

Based on the above-mentioned analysis, it is apparent that most of the generated patterns were trips to Kiyomizu Temple (S15) for both clusters 1 and 2. This is consistent with the reality that most tourists travelling to the Higashiyama area visit the Kiyomizu Temple. In addition, cluster 2, which includes overnight visitors, has a longer trip pattern and more variety of movement patterns than cluster 1 (same-day tourists). These patterns include the two main roads to Kiyomizu Temple, one from sensor 20 to sensor 15 and the other from sensor 19 to sensor 15, which is consistent with the geographic structure. Moreover, our results indicate that cluster 2 wanted to visit places other than Kiyomizu Temple. They visited more locations than cluster 1, which is also justified by the clustering result.

6.5.4 Extracting patterns that are not apparent

To check whether the SPM method can be used to extract more patterns that are not apparent to us. I selected patterns with support lower than the last pattern in Table 6.12 (0.0331) for cluster 1 and patterns with support lower than the last pattern in Table 6.13 (0.0850) for cluster 2. Tables 6.14 and 6.15 summarised these patterns.

Table 6.14 Non-obvious patterns of cluster 1.

Patterns	Support	Confidence	Lift
<{"S8"},{"S9"}> => <{"S10"}>	0.0196	0.62	2.03
<{"S11"},{"S12"}> => <{"S13"}>	0.0122	0.53	3.28
<{"S20"},{"S10"}> => <{"S6"}>	0.0106	0.48	1.77
<{"S13"},{"S12"}> => <{"S11"}>	0.0095	0.51	3.95
<{"S13"},{"S9"}> => <{"S10"}>	0.0095	0.55	1.81
<{"S2"},{"S3"}> => <{"S3"}>	0.0080	0.50	6.14
<{"S8"},{"S13"},{"S14"}> => <{"S14"}>	0.0080	0.65	1.87
<{"S20"},{"S19"},{"S10"}> => <{"S6"}>	0.0074	0.56	2.08
<{"S13"},{"S4"}> => <{"S2"}>	0.0073	0.45	3.28
<{"S14"},{"S12"}> => <{"S11"}>	0.0072	0.44	3.43
<{"S13"},{"S8"},{"S9"}> => <{"S10"}>	0.0064	0.64	2.12
<{"S1"},{"S6"},{"S10"},{"S11"}> => <{"S19"}>	0.0028	0.59	2.14

Table 6.15 Non-obvious patterns of cluster 2.

Patterns	Support	Confidence	Lift
<{"S6"},{"S10"},{"S15"},{"S15"},{"S14"},{"S18"},{"S19"},{"S10"}> => <{"S6"}>	0.0037	0.73	1.70
<{"S11"},{"S12"},{"S13"},{"S14"},{"S14"},{"S14"},{"S14"},{"S13"}> => <{"S8"}>	0.0037	0.60	1.94
<{"S10"},{"S18"},{"S14"},{"S15"},{"S15"},{"S14"},{"S11"}> => <{"S6"}>	0.0037	0.57	1.34
<{"S2"},{"S2"},{"S9"},{"S12"},{"S12"},{"S12"},{"S14"}> => <{"S14"}>	0.0037	0.89	1.19
<{"S1"},{"S6"},{"S14"},{"S15"},{"S15"},{"S14"},{"S10"}> => <{"S6"}>	0.0037	0.66	1.55
<{"S9"},{"S11"},{"S12"},{"S12"}> => <{"S13"}>	0.0037	0.58	1.51
<{"S6"},{"S10"},{"S19"},{"S14"},{"S14"},{"S15"},{"S14"},{"S14"},{"S19"}> => <{"S19"}>	0.0037	0.73	1.51
<{"S12"},{"S11"},{"S11"},{"S6"},{"S6"}> => <{"S1"}>	0.0037	0.52	1.83
<{"S15"},{"S14"},{"S13"},{"S8"},{"S12"}> => <{"S11"}>	0.0037	0.60	2.13
<{"S15"},{"S14"},{"S14"},{"S14"},{"S14"},{"S13"},{"S12"},{"S10"}> => <{"S6"}>	0.0037	0.54	1.28
<{"S6"},{"S8"},{"S8"},{"S13"},{"S13"},{"S14"},{"S14"},{"S15"}> => <{"S15"}>	0.0037	0.90	1.25
<{"S10"},{"S11"},{"S13"},{"S10"}> => <{"S6"}>	0.0026	0.66	1.54
<{"S14"},{"S7"},{"S7"},{"S13"},{"S8"},{"S9"},{"S10"}> => <{"S6"}>	0.0026	0.60	1.40
<{"S7"},{"S13"},{"S8"},{"S8"},{"S4"},{"S4"}> => <{"S2"}>	0.0026	0.66	2.48
<{"S5"},{"S5"},{"S5"},{"S9"},{"S13"},{"S14"}> => <{"S14"}>	0.0026	0.91	1.22
<{"S19"},{"S14"},{"S14"},{"S15"},{"S14"},{"S14"},{"S8"},{"S8"},{"S4"}> => <{"S4"}>	0.0026	0.63	3.89

Figures 6.13 and 6.14 help to visualise the extracted patterns. Compared with Figures 6.11 and 6.12, it can be identified that tourists also visited the area where sensors 8 to 13 were installed other than Kiyomizu Temple. This area is popular with traditional buildings and there are many food/cafes and shops in this quaint region. Moreover, there is a good photo angle on the street between sensors 11 and 12, with Hokan-Ji Temple (near sensor 12) as the background, as Figure 6.15 shows.

Figure 6.13 Non-obvious patterns of cluster 1

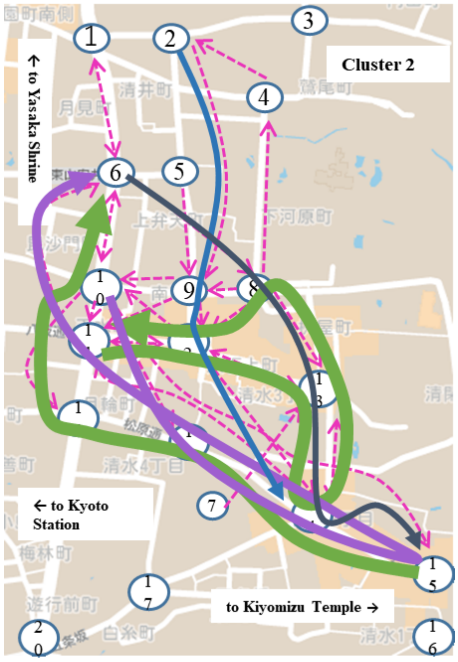

Figure 6.14 Non-obvious patterns of cluster 2

Figure 6.15 A screenshot of the street between sensors 11 and 12 from google street view

6.6 **Conclusions and future research**

This chapter used Wi-Fi packet sensors to identify the type of smart device users and the most frequently used routes by tourists in Kyoto city. The MAC address from enabled wireless communication devices was key to tracing traveller movements in the targeted area. Passive probe data from enabled wireless communication devices were integrated with clustering analysis and the SPM method. Clustering analysis was used to identify the type of smart device users, and the CSPADE algorithm was used to extract frequent sequential patterns related to destinations visited. As a result, the smart device users were categorised into four groups: same-day visitors, overnight visitors, commuters, and residents. Moreover, it was found that the frequent trip patterns of tourists involved trips to Kiyomizu Temple, which means that most of the tourists travelling to the Higashiyama area will visit the Kiyomizu Temple. These findings are useful for developing strategies for destination management, tourism management, and disaster evacuation planning in the Higashiyama area. For example, we can match the trajectory information of tourists with maps to discover their POIs. Through analysing the volume and stay time at a POI, we can know their interest and preference of the sightseeing spot. Then we can give them a label based on their preference and deliver advertisement content that they will like, thus promoting more consumption of tourists more accurately. These analytical methods may be applicable to other tourist destinations and pedestrian flow studies, such as passenger flow inside a transfer station. I expect that we can extract frequent patterns even when they are not apparent to us. Based on the SPM results, it can identify the significant location for the further development plan and estimate the next destination of the pedestrians. This kind of information will positively affect business stores along the street and save time for drivers who want to pass through the area. In addition, providing such information to some smart device applications would enable tourists to arrange their visiting destinations in line with less busy periods. This study provides novel information to the authorities and organisations involved in traffic and tourism management. The findings will allow them to make sound decisions on policies and plans to relieve congestion, as well as to improve the quality of life of residents and the tourism industry's environment, and enhance sustainable tourism development in the area. This technology can also help in monitoring the pedestrian's travel changes before and after the COVID-19 pandemic and emergency declaration. For example, it can observe the volume and stay time of pedestrians at a public place like a mall or restaurant and can also observe the use of public transportation and cross-city mobility. Such analysis will be done in further study.

References

Agrawal, R., Srikant, R. (1995). Mining sequential patterns. In Proceedings of the 11th International Conference on Data Engineering, Taipei, Taiwan, 3–14.

Aseervatham, S., Osmani, A. (2005). Mining short sequential patterns for hepatitis type detection. Presented at the ECML/PKDD 2005 - Discovery Challenge Workshop, Portugal, 6.

Batal, I., Valizadegan, H., Cooper, G. F., Hauskrecht, M. (2011). A pattern mining approach for classifying multivariate temporal data. In 2011 IEEE International Conference on Bioinformatics and Biomedicine, 358–365.

Bayarma A., Kitamura R., Susilo Y. O. (2007). Recurrence of daily travel patterns: Stochastic process approach to multiday travel behavior. Transportation Research Record, 2021(1), 55–63.

Bermingham, L., Lee, I. (2014). Spatio-temporal sequential pattern mining for tourism sciences. Procedia Computer Science, 29, 379–389.

Bin, C., Gu, T., Sun, Y., Chang, L. (2019). A personalized POI route recommendation system based on heterogeneous tourism data and sequential pattern mining. Multimedia Tools and Applications, 78(24), 35135–35156.

Buchta C., Hahsler M., Diaz D. (2020). ArulesSequences: Mining frequent sequences. R package version 0.2-25 (https://CRAN.R-project.org/package=arulesSequences).

Dua, D., Taniskidou, E. (2017). UCI Machine Learning Repository. The University of California, School of Information and Computer Science, Irvine (https://archive.ics.uci.edu/ml/datasets/Taxi+Service+Trajectory+-+Prediction+Challenge,+ECML+PKDD+2015).

Exarchos, T. P., Papaloukas, C., Lampros, C., Fotiadis, D. I. (2006). Protein classification using sequential pattern mining. In 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, 5814–5817.

Fournier-Viger, P., Gueniche T., Tseng, V. S. (2012). Using partially-ordered sequential rules to generate more accurate sequence prediction. In: International Conference on Advanced Data Mining and Applications, 431–442, Springer, Berlin, Heidelberg.

Fournier-Viger, P., Roger, N., Engelbert, M. N. (2008). A knowledge discovery framework for learning task models from user interactions in intelligent tutoring systems. In: Mexican International Conference on Artificial Intelligence, 765–778, Springer, Berlin, Heidelberg.

Ibrahim, R., Shafiq, M. O. (2019). Detecting taxi movements using random swap clustering and sequential pattern mining. Journal of Big Data, 6(1), 39.

Izakian, Z., Mesgari, M. S., Abraham, A. (2016). Automated clustering of trajectory data using a particle swarm optimization. Computers, Environment and Urban Systems, 55, 55–65.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. Pattern Recognition Letters, 31(8), 651–666.

Jin, H., Chen, J., He, H., Williams, G. J., Kelman, C., O'Keefe, C. M. (2008). Mining unexpected temporal associations: Applications in detecting adverse drug reactions. IEEE Transactions on Information Technology in Biomedicine, 12(4), 488–500.

Julea, A., Méger, N., Trouvé, E., Bolon, P. (2008). On extracting evolutions from satellite image time series. In IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium, 5, V-228.

McAullay, D., Williams, G., Chen, J., Jin, H., He, H., Sparks, R., Kelman, C. (2005). A delivery framework for health data mining and analytics. In Proceedings of the 28th Australasian Conference on Computer Science, 381–387.

Mooney, C. H., Roddick, J. F. (2013). Sequential pattern mining–approaches and algorithms. ACM Computing Surveys (CSUR), 45(2), 1–39.

Norén, G. N., Bate, A., Hopstadius, J., Star, K., Edwards, I. R. (2008). Temporal pattern discovery for trends and transient effects: Its application to patient records. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 963–971.

Pokou, Y. J. M., Fournier-Viger, P., Moghrabi, C. (2016). Authorship attribution using small sets of frequent part-of-speech skip-grams. In Flairs Conference, 86–91.

Reps, J., Garibaldi, J. M., Aickelin, U., Soria, D., Gibson, J. E., Hubbard, R. B. (2012). Discovering sequential patterns in a UK general practice database. In Proceedings of the 2012 IEEE–EMBS International Conference on Biomedical and Health Informatics, 960–963.

Schweizer, D., Zehnder, M., Wache, H., Witschel, H. F., Zanatta, D., Rodriguez, M. (2015). Using consumer behavior data to reduce energy consumption in smart homes: Applying machine learning to save energy

without lowering comfort of inhabitants. In 2015 IEEE 14th International Conference on Machine Learning and Applications, 1123–1129.
Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. IOP Conference Series: Materials Science and Engineering, 336(1), 012017.
Wang, J., Han, J., Li, C. (2007). Frequent closed sequence mining without candidate maintenance. IEEE Transactions on Knowledge and Data Engineering, 19(8), 1042–1056.
Wright, A. P., Wright, A. T., McCoy, A. B., Sittig, D. F. (2015). The use of sequential pattern mining to predict next prescribed medications. Journal of Biomedical Informatics, 53, 73–80.
Zaki, M. J. (2000). Sequence mining in categorical domains: Incorporating constraints. In Proceedings of the 9th International Conference on Information and Knowledge Management, 422–429.
Zaki, M. J. (2001). SPADE: An efficient algorithm for mining frequent sequences. Machine Learning, 42(1–2), 31–60.
Ziebarth, S., Chounta, I. A., Hoppe, H. U. (2015). Resource access patterns in exam preparation activities. In Design for Teaching and Learning in a Networked World, 497–502.

Chapter 7: Conclusion, key contributions and future work

This research mainly aims to gain a better understanding on analysing pedestrian behaviour utilising Wi-Fi packet sensor data. In this final chapter, we summarise outcomes of the results in section 7.1, and key contributions related to science and society in section 7.2. We latter discuss the opportunities for future studies in section 7.3.

7.1 Summary of thesis

Chapter 2 summarised various data collection methods for pedestrian behaviour analysis. Compared with other pedestrian data collection methods, Wi-Fi packet sensor based data collection system does have advantages such as convenience to handle the data, and an ability of obtaining long-term observations data and real-time observation data at low cost. However, the Wi-Fi packet sensor observes the smart devices carried by users instead of observing the pedestrians directly (manual count or video-based methods), thus the correlation between the sensor observation and real pedestrian data should be clarified.

Chapter 3 can be regarded as a fundamental analysis based on the question 'What factors affect the data collection of the Wi-Fi packet sensor and whether it is suitable for collecting pedestrian data utilizing the Wi-Fi packet sensor?'

It is concluded that both the installation height and type of sensor can influence the detection results. Furthermore, it identified that the Wi-Fi packet sensor is more suitable to observe pedestrians than vehicles through comparing the approximation functions for modelling vehicle and pedestrian flow. I also explored whether I could categorise our campus Wi-Fi packet sensor data to elucidate pedestrian behavioural patterns and checked the spatial and temporal variation of pedestrian flow. As a result, I could characterise the campus data into three groups namely: undergraduate students, stationary devices and graduate students. Moreover, a Sankey diagram was used to show the spatial and temporal variation of pedestrian flow.

Chapter 4 deals with the question 'How the installation conditions influence the observation result of the sensor and how the pedestrian count can be estimated?'

This chapter quantified the influence factors of the Wi-Fi packet sensor based on the two-sensors method. The Wi-Fi packet sensor does not observe pedestrians directly but it detects the device carried by the people. It is therefore needed to understand the correlation between the real pedestrian count and sensor observations. Based on the analysis in this chapter, the correlation between the Wi-Fi packet sensor observations and real pedestrian count may fit well with the exponential function was identified and how the installation height and environmental conditions influence the data collection of the sensor was estimated.

Chapter 5 addressed the question ‘Is it possible to evaluate the crowding level of visitors of a street?’

Based on the quantified installation parameters of the sensors in chapter 4. This chapter attempted to evaluate the PLoS of any segment. Based on the installation height and surrounding condition information, it is possible to estimate the pedestrian flow count of a segment and then estimate the PLoS level with the sidewalk width. The links that has high LoS level are consistent with the most frequent trip patterns in chapter 6.

Chapter 6 addressed the question ‘Whether it is possible to analyse the behaviour of a specific group (tourist) with this anonymization data?’

This chapter tried to extract attributes of the smart device users from the anonymous observations collected from a tourist area. As a result, I could characterise the smart device users into four groups: same-day visitor, overnight visitor, commuters, and stationary devices and residents. Moreover, it was found that the most frequent trip patterns of tourists matched our expectation and I concluded that the proposed method can extract ‘golden routes’ of other public places.

7.2 **Key contributions**

As one of the important parts of traffic, pedestrian research has always attracted the attention of researchers. The development of proposed technologies provides new chances to collect pedestrian data for the research of pedestrian behaviour. In this section, scientific and social contributions of this thesis are summarised as follows.

7.2.1 Scientific contributions

This thesis explored how we can extract useful information from Wi-Fi packet sensor data for studying pedestrian mobility behaviour. The scientific contributions can be summarised as follows：

1. Proposing an integrated data processing method of Wi-Fi packet sensor.

The first contribution is to arrange Wi-Fi sensing data to something that can be used for transport (or pedestrian) analysis. It verified that it is possible to detect pedestrians and track their locations by means of Wi-Fi sensing technology. Because the MAC address is unique per device, it is possible to aggregate the observation of a specific pedestrian at different places and analyse the spatial-temporal behaviour of crowds.

2. Proposing evaluation method of PLoS based on Wi-Fi packet sensor data.

Although we need to improve the accuracy of the model, we could estimate the model to evaluate the pedestrian counts quantitatively from Wi-Fi sensing data. Also based on the estimated pedestrian count, we can evaluate PLoS following to HCM manual.

3. Identifying the effectiveness of clustering analysis to infer traveller attributes.

It is confirmed that clustering analysis can infer the traveller attributes: tourist, regular commuter and so on. It is generally said that it is difficult to infer these attributes from such passive big data like Wi-Fi or GPS, but clustering analysis contribute to tackle this issues.

4. Identifying the effectiveness of Sequential Pattern Mining (SPM) to extract frequently used routes by travellers.

It also confirmed that SPM can extract the frequently used routes based on the massive trajectory data collected by the Wi-Fi packet sensor. Providing this information to intended tourists can support them arranging the destination visiting sequence.

7.2.2 Societal application

The Wi-Fi packet sensor data can be applied to many fields. Based on this research we can extend its applications to:

1. Showing importance of EBPM on tourism management

The discussion in Chapter 5 concludes the effectiveness of ‘light up’ event on pedestrian or traffic flow. This example suggests that such tourists event may influence on traffic or pedestrian conditions positively or negatively. By using Wi-Fi sensor, we can evaluate the influence of such events onto traffic quantitatively. Like this, we need to move to ‘evidence-based policy making’ (EBPM) even for tourism management.

2. Importance of marketing of travellers

Clustering analysis could easily infer travellers into groups with different interests. Some types of travellers may flow into busier areas whereas others not. Different preference can be found in different type of traveller. This fact suggests an importance of providing different information for different people.

7.3 **Future work**

In chapter4, we have survey data of only six cross-sections. The parameters need to be justified with data collected from more cross-sections. The data used in the research is collected by 20 Wi-Fi packet sensors and the study object is also a small area. It will be more significant to study pedestrian mobility behaviour in an urban area with more sensors.

Complex urban transport systems include competing modes between pedestrians and vehicles. Pedestrian flow may influence largely onto traffic especially in cities like Kyoto where lots of tourists visit, and efficient control of pedestrian flow may contribute to improve the vehicle-traffic conditions. Therefore, there is a need to understand the network flow dynamics for both vehicles and pedestrians. Combining Wi-Fi packet sensor data and other data sources such as traffic detector data to analyse the interaction between vehicle traffic and pedestrian traffic should be done in the future study.

The Wi-Fi packet sensor can also be used to:

1. Urban design and management

The Wi-Fi packet sensor data can help to capture and analyse how people move and interact with urban infrastructures. It can be used to analyse the demand of users from the temporal and spatial perspectives. The sensor can also be used to monitor the urban commuter behaviours such as their origin and destination which can be reference information for urban management and design. Moreover, this data collection method can also be applied for public security and emergency response, for example, regional heat map in public areas can provide necessary data support for social security.

2. Public transportation systems

The Wi-Fi packet sensor can also be used to monitor passenger crowding in buses, and trains, at bus stops and in railway/metro stations in real-time. The real-time crowding information can be transmitted to a smart subsystem to support the crowd control functionality through a communication infrastructure. The real-time crowding information can be used by PT operators for fast or proactive adaption of some service changes (increasing the frequency of service, reallocating vehicles from one line to another, planning alternative routes) in order to cope with spatially and temporally localized crowding situations. Such information can also be reported by means of displays installed inside vehicles or at stations, or through mobile apps to allow for safe PT usage during exceptional events outbreak, such as COVID-19.

3. Application in business

This new population sensing technology also provides opportunities in retail analysis to predict future sales. For example, customer flow statistics, identify new and old customers, length of stay in the store, regional heat map, crowd trajectory, visiting cycle and so on.

4. Support effectiveness of Smart Cities

The key concept of smart cities is to continually observe the social conditions by using recent information technology. Wi-Fi sensing can also be used for data observation. It can also be used to dynamic pedestrian behaviour management in the urban environment since the data can be obtained in real time.