USING GEO-TAGGED TWEETS FOR UNDERSTANDING ACTIVITY DISTRIBUTIONS IN KYOTO

Mingyu BI[1], Wenzhe SUN[2], Jan-Dirk SCHMÖCKER[3], Yue MA[4],
Borja MOYA-GÓMEZ[5], Satoshi NAKAO[6], Tadashi YAMADA[7]

[1] Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: bi.mingyu.75s@st.kyoto-u.ac.jp

[2] Member of JSCE, Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: wz.sun@trans.kuciv.kyoto-u.ac.jp

[3] Member of JSCE, Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: schmoecker@trans.kuciv.kyoto-u.ac.jp

[4] Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: ma.yue.55t@st.kyoto-u.ac.jp

[5] tGIS Research Group, Department of Geography, Complutense University of Madrid
(Universidad Complutense de Madrid, Madrid 28040, Spain)
E-mail: bmoyagomez@ucm.es

[6] Member of JSCE, Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: nakao@trans.kuciv.kyoto-u.ac.jp

[7] Member of JSCE, Graduate School of Management, Kyoto University
(Kyoto daigaku-katsura, Sakyo-ku, Kyoto 606-8317, Japan)
E-mail: yamada.tadashi.2x@kyoto-u.ac.jp

The emerging social networking services and the continuous improvement of their functions in the last decade are offering new ways to measure activity demand. We aim to examine the feasibility of using geo-tagged tweets to infer activity participation and its spatial and temporal distribution. We extracted “Swarm check-ins” from the tweets and calculated the number of check-ins per Point of Interest (POI). Then we attached an activity type to a POI using Foursquare place categories and aggregated activity participation per activity type. A case study is conducted in 37 tourist zones of Kyoto City using the data collected from July 2021 to December 2022. The estimated activity demands show reasonable and explainable patterns in popular tourist zones and changes according to tourism seasons and COVID-19 restrictions. We conclude that qualitative and quantitative analysis of event participation using this data source is feasible.

Key Words : *activity participation, activity demand, geo-tagged tweets, check-in, COVID-19, tourism*

1. INTRODUCTION

Travel demand is a derived demand of activity participation. Understanding patterns of activity participation in major areas of the city may be significant in many contexts, including urban planning, traffic management, and activity proposals of citizens and visitors, and may even yield a deeper understanding of the geographic evolution of urban social life.

Traditionally, it is common to conduct surveys on activities and travel over the course of a day to obtain patterns of activity participation. For example, household surveys are asking about activity types, geographic location, transport means used, and start/end time. Thus the responses include rich information about the activity and transport to and from these. Even more detailed on the activities are “time-use surveys” as for example analyzed recently in Pabilonia and Vernon (2022). However, unless large incentives are used, it is difficult to receive sufficient responses to detailed surveys as the questionnaire is cumbersome for respondents to be completed.

With the development of the Internet and smartphones, a growing number of individuals are posting all types of information on SNS (Social Networking Service), and they are sharing their real-world live status anytime and anywhere in the virtual online world. This also provides us with new ways to obtain information about people's daily activities. Geo-tagged tweets are a good example.

Enabling precise location allows users to optionally add location information to their tweets. This feature is turned off by default and users will have to select it to use it. If turned on, this will allow Twitter to collect, store and use their precise location, such as GPS information.

There have been some studies on geo-tagged tweets in the past, including studies on the analysis of activity types. As the literature review will show, the current studies analyze user activity in a very general scope and lack analysis of the characteristics of different periods. Many studies only distinguish commuting and non-commuting activities, whereas others are more detailed on activity types but with a low spatial resolution. Our study identifies several key activities including dining and sightseeing. Besides this, we study the changes in the spatial and temporal distribution of activity types, focusing on the distribution of different types of activities in specific regions and specific periods, making it easier for administrative reference to analyze the reality of activities.

We aim to find the main activity types and purposes in a certain area and the popular venues by estimating the type of activity using Twitter data. By analyzing the activity demand patterns in different seasons and time-of-day periods, we examine the feasibility of using Twitter data to analyze activity participation. Furthermore, we can then expect the results to be used to determine the difference between the "planned" and actual use of the area which then can support future urban planning. For example, based on land-use characteristics, shopping might be the main activity in an area, but if most tweets report about a particular restaurant, one can learn that this restaurant might be important for the overall attractiveness of the area.

The remainder of this paper is organized as follows. Section 2 reviews the existing literature using geo-tagged tweets. Our datasets of geo-tagged tweets and Foursquare POIs are described in Section 3 in detail. Section 4 elaborates on the methodology for tweet cleaning and activity matching, with the percentage of tweets that could be related to an activity. Section 5 showcases the temporal distribution of the key activities estimated from tweets and provides a closer investigation of the patterns of activity participation in specific areas during specific time-of-day periods. The results also illustrate the dynamics of activity demand during COVID-19 and tourism seasons. The findings of this research and future research directions are discussed in Section 6.

2. LITERATURE REVIEW

Kobayashi (2022) conducted a study on the correlation between tweets and mobility during the epidemic period of COVID-19. The study aimed to understand the changes in people's emotions during the pandemic of COVID-19 and compared them with population flow data to examine the possibility of predicting changes in human flow based on the content of tweets. The results suggest that there is a correlation between the number of people observed at transit stations and the number of emotions, especially those categorized as "joy". Furthermore, the use of emotion in tweets is shown to be useful in predicting the number of people visiting the city. High anxiety about COVID expressed in the Tweets was found to be associated with fewer tourists visiting the city in the next weeks.

García-Palomares et al. (2017) used Twitter data to analyze city dynamics over the day. Users of this social network were grouped according to city zone and time slot to analyze the daily dynamics of the city and the relationship between this and land use. Specifically, they obtained typical Twitter activity profiles based on major land uses in each district, showing how activity-related land uses are activated during the day, and performed multiple regression analyses to determine the impact of different land uses on each major period. The results suggest that Twitter data can be useful in helping to improve our understanding of the link between land use and urban dynamics.

Martin et al. (2020) examined the suitability of Twitter data for measuring post-disaster population mobility using the case of Hurricane Maria in Puerto Rico. They tracked Twitter users living in Puerto Rico by using geo-tagged tweets to check how many are displaced, the timing and destination of the users' displacement, and whether they returned. It is found that 8.3% of the resident sample relocated during the months, and several visitors to Puerto Rico were deceased. Thus, the potential usage of geo-tagged tweets to understand the mobility patterns during a disaster was confirmed.

Lee et al. (2012) proposed a method for monitoring crowd behavior in urban space using location-based tweets and extracting regional characteristics of cities. They extracted crowd behavior patterns in urban areas by analyzing crowd behavior vectors created based on geo-tagged tweets.

Noulas et al. (2011) investigated how check-ins of the Foursquare application reflected geographic user

Table 1 Details of missing periods.

Missing periods	Number of days	Missing periods	Number of days
2021/07/09 to 2021/07/28	20	2022/05/12 to 2022/05/24	13
2021/08/12 to 2021/08/23	12	2022/06/16 to 2022/06/20	5
2021/09/17 to 2021/09/22	6	2022/08/10 to 2022/08/18	9
2021/10/14 to 2021/10/21	8	2022/09/14 to 2022/09/28	15
2021/11/12 to 2021/11/18	7	2022/10/12 to 2022/10/19	8
2022/02/10 to 2022/02/14	5	2022/11/09 to 2022/11/15	7
2022/03/10 to 2022/03/25	16	Total	131

activity patterns. They analyzed user check-in dynamics, demonstrating how it reveals meaningful spatio-temporal patterns and offers the opportunity to study both user mobility and urban spaces. It is found that activities recorded by these Foursquare check-ins vary within a day and a week, with meaningful patterns closely related to human activity from a temporal and spatial point of view.

Overall, we conclude that the available literature confirms that Tweets appear to be useful to understand mobility patterns, in particular during exceptional circumstances. Few studies have explored the spatial and the seasonal profile of the Tweets from a city which is the objective of the current study.

3. DATA DESCRIPTION

3.1 Tweet data

We collected geo-tagged tweets posted in Japan from Jul. 1st, 2021 to Dec. 4th, 2022 (522 days); although we failed to collect tweets on some days because of computer issues. In total on 131 days the data collection was not completed and we omitted such days. The detailed periods are shown in **Table 1**. The collected data set consists of 14,164,489 geo-tagged tweets collected from all around Japan. Among these, 270,132 tweets were posted in Kyoto City and subsequently used for our analysis.

The information we have collected includes tweet ID, coordinates (longitude and latitude), general place (administrative region, city, or neighborhood), place ID, tweet text, and user information covering the user ID and when the user signed up for the Twitter account. Furthermore, sometimes, the text contains one or more URLs. This is the case if users add a URL to their tweet content, or it could be a shared tweet from a third party. Hence, we can obtain further information through the URL, for example, "Swarm posts" could be found through the URL for exploring which location the person checked into. Swarm, which is also called "Foursquare Swarm", is an app that is used specifically to share one's location with friends and to create a history of places visited. The slogan of Swarm is "remember everywhere" (swarmapp.com).

Twitter has interfaces with applications such as Swarm and Instagram. Accordingly, tweets can be divided into 2 types depending on the methods used to post as illustrated in **Figure 1**: "Shared tweets" and "direct posting" tweets. Shared tweets are those which are posted to third-party applications or websites such as Swarm and Instagram and then "forwarded" to Twitter. In other words, once the user allows the third-party application to share the posts to Twitter, future posts appear on their Twitter account as well. Interestingly, we found over 80% of geo-tagged tweets are shared from Swarm or Instagram, and only a few users prefer to directly tweet with their locations.

Tweets can also be grouped by location labels: general location labels or specific place name labels besides business, landmark, or point of interest. For example, when users tweet, they can only tap the location marker to add an administrative region location label such as 'Kyoto-Shi, Sakyo-Ku'. Also, they can choose a specific place name as their location labels such as 'Kyoto University Yoshida Campus'. These places are sourced from Foursquare using the same location labels as the shared tweets from Swarm. Therefore, a fairly high proportion of geo-tagged tweets can be associated with the Foursquare Points of Interest (POIs).

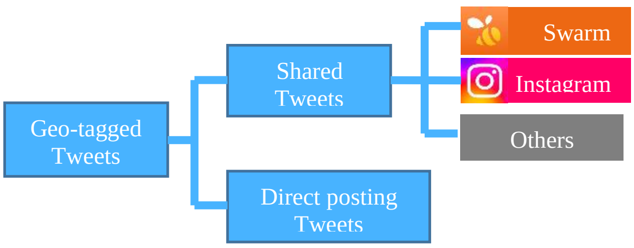


Fig.1 Types of geo-tagged tweets.

3.2 Foursquare POI data

We access the Foursquare API to retrieve the Foursquare POIs from which a Swarm user can choose to "check-in". The information of a POI mainly includes its name, coordinates, formatted address, postcode, region, country, Foursquare ID, street name, parent category, and specific location ID and type. **Figure 2** shows an example of POIs in central Kyoto. As can be seen, the density of points in central Kyoto is fairly high. Foursquare Places include a hierarchical taxonomy of categories from which each POI record is classified. For example, the location type of a Ramen restaurant is under the parent category of "Dining and Drinking". The ten parent categories obtained from FPO (2023) are: Arts and Entertainment, Business and Professional Services, Community and Government, Dining and Drinking, Event, Health and Medicine, Landmarks and Outdoors, Retail, Sports and Recreation, and Travel and Transportation

In this research, a parent category is regarded as an activity type. In the data we collected in Kyoto City, there were no check-ins under the "Event" category. Further, "Sports and Recreation", "Health and Medicine", and "Business and Professional Services" include too few check-ins for a meaningful further analysis. Thus, we determined to focus on check-ins under the remaining six categories that are also shown in **Figure 2**.

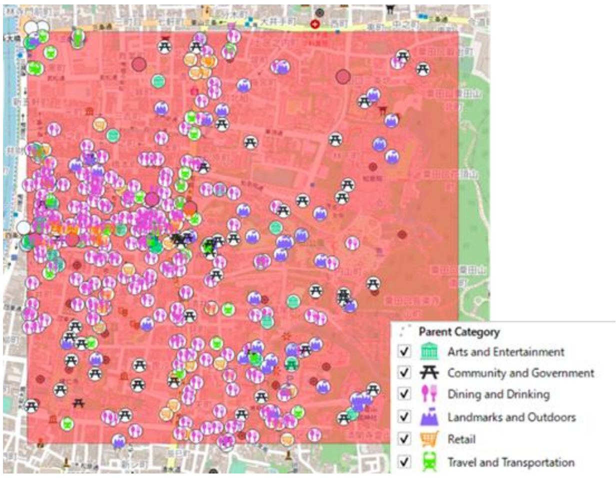


Fig.2 Illustration of Foursquare POIs in the Kyoto, Gion area.

Under every of these parent categories, there are plenty of location types. The name and total check-in number of the main location types in each parent category are shown in **Table 2**. Note that some temples and shrines are divided into the category of Community and Government, while others are in the Landmarks and Outdoors. This indicates the multiple functions of a religious place, which has a meaning for both the local community and tourism activities. To note is that, for example, Ramen restaurants obtain a very high number of tweets. This will be partly because of the bias in Twitter and Swarm users towards younger people who are more often visiting these fast-food restaurants. Secondly, it also reflects that Ramen is an important point of discussion among the Japanese. That is, conversations about which Ramen shops are good and recommendable are a common topic among Japanese. Hence, having checked-in at a good Ramen restaurant might attract responses from friends. At the same time, not only "cool locations" are observed frequently in our database. As **Table 2** shows also being at one of the many convenience stores is frequently posted. Such locations might be used for appointments with friends and hence the tweet lets the others know that one has arrived. We suspect a similar reason applies to the most popular tweet category "rail station". A large number of these Tweets will be for appointment purposes or generally for letting others know that one has arrived in Kyoto. Rail stations such as "Kyoto station" or "Karasuma station" are easy-to-understand locations for those who follow the user's tweet account without the need to look up where this location is. Overall, we conclude that the Tweets certainly have biases but also show a wide distribution of activities and locations.

Table 2 Activity grouping, location types, and check-in numbers in our data.

Parent activity category	Location type	Check-in number
Arts and Entertainments	Concert Hall	454
	Movie Theater	321
	Karaoke Box	243
	Public Art	216
	Others	1106
Dining and Drinking	Ramen Restaurant	2171
	Cafe´	1192
	Sake Bar	622
	Bar	586
	Coffee Shop	461
	Japanese Restaurant	399
	Chinese Restaurant	383
	Burger Joint	311
	Bakery	282
	Others	3863
Community and Government	Buddhist Temple	1755
	Shrine	1602
	Temple	388
	Spiritual Center	261
	Post Office	114
	College and University	71
	Public and Social Service	54
	College Academic Building	32
	Library	32
	Government Building	31
	Others	190
Landmarks and	Historic and Protected Site	2589
	Bridge	782

Parent activity category	Location type	Check-in number
Outdoors	Monument	544
	Park	336
	Waterfront	218
	Garden	214
	Bathing Area	128
	Hot Spring	104
	Scenic Lookout	99
	Others	387
Retail	Convenience Store	1046
	Electronics Store	859
	Shopping Mall	784
	Department Store	463
	Grocery Store / Supermarket	384
	Golf Store	217
	Clothing Store	159
	Bookstore	138
	Candy Store	116
	Flower Store	114
	Others	1405
Travel and Transportation	Rail Station	15844
	Bus Stop	1823
	Metro Station	1234
	Hotel	517
	Public Transportation	515
	Bus Station	258
	Parking	226
	Tram Station	202
	Platform	145
	Rest Area	101
	Others	411

4. METHODOLOGY

4.1 Cleaning

First, we detected "active users" with very frequent posts and attempted to exclude some of them. The large number of tweets created by an active user, often from the same location, would lead the statistical analysis to be biased towards their tweets. A criterion to determine if they are active or not is whether they have posted 100 or more tweets during our analysis period, which equates to more than one tweet every four days. Then, for each user we consider active, we counted how often they tweeted at each location. Active users can be divided into three groups according to the characteristics of the distribution of the number of tweets by location. 1) The tweets are all concentrated in a very few or even one posting location; 2) Multiple posting locations, but with a very high number of tweets from a few locations; 3) Multiple posting locations, a relatively even number of tweets per posting location. We further inspected the tweet content and whether the tweets are shared or direct ones. This leads to the grouping and classification of "user types" shown in **Table 3**.

Table 3 Data distribution characteristics.

Data distribution characteristics	Main tweet types	User types
The posted tweets are all concentrated in very few or even one posting location	Direct posting, From Instagram	Business user
Multiple posting locations, with a very high number of tweets from a few locations	Direct posting, From Instagram From Swarm	Business user, Personal user, Personal business hybrid users
Multiple posting locations, relatively even number of tweets per posting location	Direct posting, From Instagram, From Swarm	Personal user

We erased all tweets in 1) and kept all tweets in 3). Since the case of 2) contains multiple user types, we manually checked whether these are personal users. If it is a personal user, we will keep the tweets of the user, but if it is likely a business user or a personal and business hybrid user, we will eliminate the tweets from the database.

For users with less than 100 tweets, it is impractical to depend on manual judgment due to the huge number of tweets. Therefore, it is difficult to distinguish all business users from individual users only by the distribution of the number of tweets on the posting location. To note is that while shared tweets from Instagram and direct posting Tweets are posted by both business and personal users, shared tweets from Swarm are almost exclusively by personal users.

Moreover, while manually checking the content of the tweets, we found that there are in some cases multiple tweets posted by the same user within a short time at the same place. Since one of the important roles of geo-tagged tweets is to let someone else know the user's location when they're going to meet each other, users possibly tweet more than once to let people know where they are. Thus, we delete repeated tweets except for the first one if one user keeps tweeting at the same place within one hour.

4.2 Matching

4.2.1 Perfect matching

After cleaning, the next step is to associate each tweet with a location and an activity type. We aimed to obtain as much information as possible from the tweets about the place, time, and content of the event.

The first type of matching, which we call "perfect matching", refers to matching Swarm check-ins with Foursquare POIs. Each tweet shared from a third-party application or website contains a URL in the text, and we extracted the Foursquare place IDs of the locations of each Swarm check-in through a web crawler. Using these IDs, we utilized the Foursquare API to collect specific information about the POI including the name of the place, location, and category as introduced in 3.2.

Thus, given that the venues and parent categories shown in **Table 2** are provided as part of the tweet information, it is expected that these tweets can reveal, at least to some degree, the activity participation behind the post. We note that it is only to some degree the activity, since, as discussed, some POIs such as station, might be chosen by the person not for the purpose of engaging in an activity at that place but because it might be an easily recognizable location.

Since the Swarm tweets form a large proportion of the total tweets, this allows us to infer the most popular POIs in a certain area or a city during a data collection period. By aggregating the total number of check-ins of the POIs in the same Foursquare's parent category for specific time periods, we obtained the spatio-temporal demand distribution of each type.

4.2.2 Fuzzy matching

Some of the Foursquare IDs were invalid and did not match any of the Foursquare API POIs, thus, we extracted their posting coordinates and conducted a "fuzzy matching" for these Tweets together with other types of tweets for which only location coordinates were available.

Fuzzy matching we define as matching only the general area of activity from these tweets but not the activity type. To be more specific, we do not infer the detailed place and type of activity from these tweets but use these tweets only for the count of the total number of tweets from each area. The total number of tweets that have been perfectly matched in a certain region is summed up to get the total number of users presumed to conduct activities in that region as a reference.

5. KYOTO CITY CASE STUDY

5.1 Tourist zone data

Kyoto city is a famous tourist city for Japanese as well as foreign travelers. It is further an interesting case study for spatial analysis of tweet distribution since it consists of multiple, distinct touristic areas distributed over a fairly large area. Also, people tend to share their location via tweets more for unusual events, e.g. sightseeing, rather than for regular activities. For these reasons, we determined to the geo-tagged tweets posted inside 37 tourism zones to analyze the activities of people and the use of facilities in these areas.

These areas are defined and used in a survey of tourist movements in Kyoto city to explore tourist preferences, conducted in November 2006 (Urban Planning Bureau of Kyoto City, 2006). Shen et al. (2022) also used this survey recently acknowledging that it is outdated but noting: "While this survey is quite old, we note that no major changes have occurred in Kyoto's tourist attractions or its transportation network over the past decade. However, we acknowledge that decisions about which areas to visit and how to travel may be different today due to the increasing number of mobile phone applications that offer advice to tourists." Our study hopes to further contribute to touristic activities in these areas without conducting an additional survey. We believe that using these areas can better leverage the characteristics of geo-tagged tweet data. The name, index, and location of zones are shown in **Table 4** and **Figure 4**.

Of all the tweets we collected, 95,781 tweets were posted in 37 Tourist Zones, accounting for about 44% of all tweets in Kyoto City. The flowchart in **Figure 3** illustrates the tweet cleaning and matching methodology explained in Section 3 with the data processing results. After cleaning up all active business users and repeated tweets there are 86,380 tweets left, which we call the "cleaned tweets data set". After checking URLs, we found 63,371 Swarm check-ins, posted from 5,412 POIs. Among them, there are 53,244 check-ins from 4,849 POIs with a valid Foursquare ID and 10,127 check-ins from POIs with an invalid Foursquare ID. Hence, for a total of 62% of our cleaned tweets data set we can find the specific check-in venue, and thus can be perfectly matched and the activity inferred.

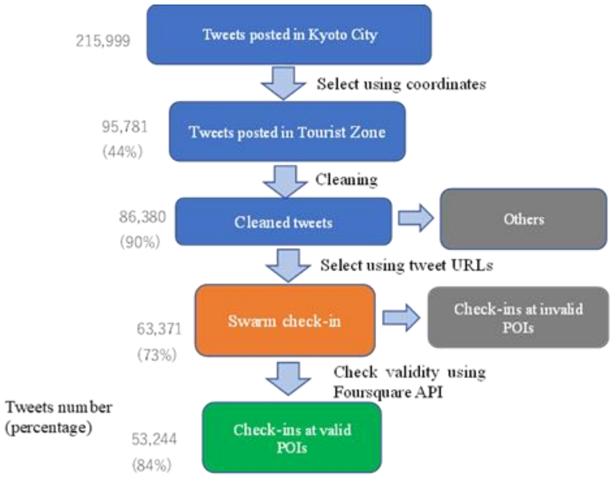


Fig.3 Flowchart of data processing.

Table 4 Index and name of the tourist zones in Kyoto.

Index	Name
1	Ohara/Yase
2	Kurama Area
3	Takaragaike
4	Kamigamo Shrine
5	Takao Yama
6	Shugakuin/Shisendo
7	Koetsu-ji Temple
8	Kitayama Dori
9	Daitokuji Temple Area
10	Kinkaku-ji Temple
11	Shimogamo Shrine
12	Kitano Tenmangu
13	Kinugasa/Omuro
14	Sagano Area
15	Ginkaku-ji Temple
16	The Path of Philosophy
17	Heian Jingu Shrine
18	Kyoto Imperial Palace
19	Hanazono Area
20	Nijo Castle Area
21	Nijo Station Vicinity
22	Uzumasa Area
23	Arashiyama Area
24	Gion Area
25	Kawaramachi
26	Matsuo Taisha Are
27	Kiyomizu-dera
28	Sanjusangendo
29	Kyoto Station Vicinity
30	Katsura Imperial Villa
31	Tofukuji Temple Area
32	Toji Temple Area
33	Fushimi Inari Shrine
34	Daigoji Temple Area
35	Jonan-gu Shrine Area
36	Fushimi Area
37	Keihoku Direction

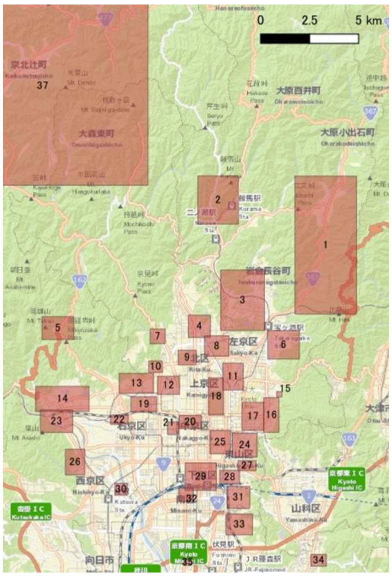


Fig.4 Tourist zones in Kyoto city, adapted from Shen et al. (2022).

5.2 Results for Kyoto city

5.2.1 All check-ins in all zones

The 24-hour distribution of the daily average of check-in activity of the six activity types on work-days and weekends during the whole data collection period is shown in **Figure 5**.

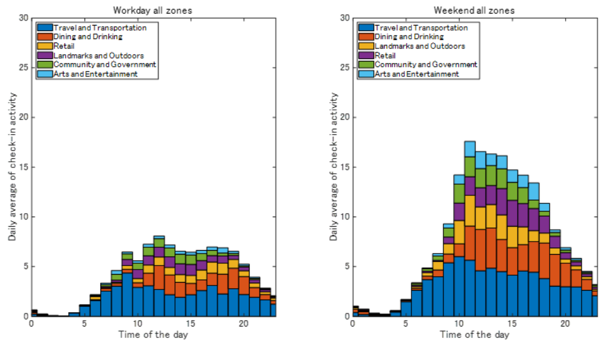


Fig.5 24h distribution of all check-ins.

We observe from **Figure 5** that the total amount of daily average activity participation is significantly higher on weekends and holidays than on workdays, indicating that tweet activities are more often for lei-sure activities. Whether workdays or weekends, the two most popular types of activities are “Travel and Transportation” and “Dining and Drinking”, and the two least popular activity types are “Arts and Enter-tainment” and “Community and Government”. We find that there is more activity participation in “Re-tail” than in “Landmarks and Outdoors” on work-days, while the ranking reverses on weekends.

On workdays, “Travel and Transportation” shows two peak values at 9:00 and 17:00, which are the times when people commute to work and home from work. On weekends, the peak of travel occurs at 10:00, after which the number is maintained rela-tively steadily until the late hours. This trend is rea-sonable because people go out later on days off than on weekdays and stay out longer.

The peaks of Dining and Drinking occur at 12:00 to 14:00 and 18:00 to 20:00 on workdays, while 11:00 to 14:00 and 17:00 to 20:00 are the hours with most tweets on weekends. To note is that 14:00 to 17:00 also remain relatively high amounts of activity participation in “Dining and Drinking”, especially on weekends. This might result from more visits to cafés on weekends.

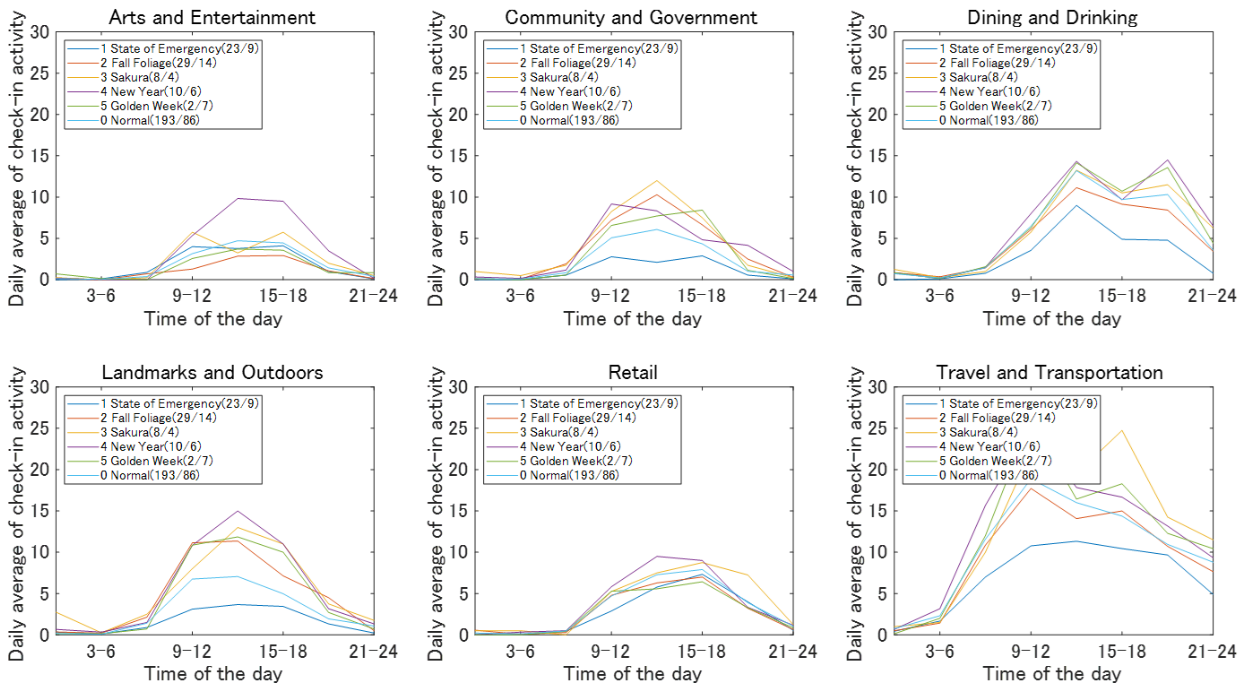
5.2.2 Check-in differences during touristic seasons and COVID restrictions

It can be expected that the 24-hour distribution of check-in activities in different periods over the year of data collection shows a difference. For one Kyoto has two key tourist seasons. During the cherry blossom in late March and during autumn foliage the tourist number markedly increases. Secondly, our data collection has been carried out during the COVID period with specific periods where social activities were discouraged. To account for this, five special periods are distinguished and compared to a "normal season" including all other days. We note that the COVID-19 restriction period is the fourth Declaration of the State of Emergency, which did not have as large an impact on activities as the initial period. The fourth declaration was not a compulsory lockdown policy though, commercial venues such as restaurants and bars are asked to close before 8 pm. Details of the periods are shown in **Table 5**. Since the total number of check-ins in every period is not large enough to further conduct an hourly analysis, we combine three hours rather than every hour to show the 24-hour tweet activity distribution.

Figure 6 shows the daily check-in activity demand in the aforementioned periods. The total number of check-ins during the State of Emergency and with it the number of tweets in each category are relatively low. In particular "Community and Government" and "Landmarks and Outdoors" had the most significant decrease in volume relative to the other periods. "Dining and Drinking" also markedly decreases whereas "Retail and Art and Entertainment" experienced the smallest decrease in volume. This seems reasonable given that retail activities are often carried

Table 5 Details of the distinguished seasonal COVID periods.

Period name			Start date and end date	Total number of days (Number of workdays/weekends)	Total number of check-ins
Special Periods	1)	State of Emergency	2021/08/20~2021/09/30	32(23/9)	4846
	2)	Fall Foliage Season	2021/11/10~2021/12/10, 2022/11/10~2022/12/04	43(29/14)	11047
	3)	New Year Holiday	2021/12/25~2022/01/05	12(8/4)	3949
	4)	Sakura Season	2022/03/20~2022/04/10	16(10/6)	4626
	5)	Golden Week	2022/04/29~2022/05/07	9(2/7)	3340
Normal season			Other days	279(193/86)	56204



* The numbers in brackets refer to the numbers of workdays and weekends (including holidays) in the associated period.

Fig.6 Check-in activity in different periods.

out individually and are not as restricted during the State of Emergency. Also, explainable is that the second peak of “Dining and Drinking”, the dinner peak, is noticeably lower than the lunch peak during the period of State of Emergency when restaurants were encouraged to close early. The difference between the two peaks in the other periods is insignificant or the dinner peak is even higher.

During autumn foliage and the “Sakura season”, there are more overall tweets per day and higher peaks in the “Community and Government”, “Landmarks and Outdoors” categories compared to the other periods. These two categories include shrines, temples, and parks where Kyoto citizens as well as tourists enjoy cherry blossoms and maple leaves. Besides, Golden Week is a one-week holiday from the end of April to the start of May every year in Japan. It is also a high season for tourism, but it shows a smaller share of “Community and Government” and “Landmarks and Outdoors”, which may be due to multiple purposes of travel other than sightseeing during that period. In conclusion, we suggest this discussion shows that the tweets can pick up major changes in activity participation, provide further insights than mobility data alone and that the data are useful to understand trends during the course of a year.

5.2.3 Check-ins in different areas of Kyoto

In this last analysis section, we pick up a number of areas with sufficient Tweet data to illustrate how the activities vary on this local scale. Here we will show three areas as examples. Details of these three areas are shown in **Table 7** referring to the areas shown in **Table 4** and **Figure 4**.

Table 7 Details, main activity type, and popular venues of areas.

Area No.	Zone [Index in Table 4]	Main activity type (workday)	Main activity type (weekend)	Popular venues
1	Kyoto Station Vicinity [29]	Travel and Transportation	Travel and Transportation	Kyoto Station, JR Kyoto Station Shinkansen, Kyoto-Yodobashi
2	Kawaramachi [25]	Dining and Drinking	Dining and Drinking	Rokakku-do, Movix Kyoto, Round-1
3	Gion Area [24]	Travel and Transportation	Dining and Drinking	Gion B bus stop, Sanjo Station, Yasaka Shrine

The subsequent figures illustrate that the main activity type varies from area to area due to its geographical location and land use attributes. Also, the main types of activities in Area No.3 are different on weekends and workdays. The percentage of activity types presented by check-in also differs among the periods. To illustrate the differences we calculated the 24-hour distribution for the entire period and the average daily activity share for each of the six periods introduced in **Table 5** separately by region in **Figures 7** to **12**.

According to **Figure 7**, the total number of check-ins is high until late hours in the Kyoto station area. Since it is the main transport station it is not surprising that especially the “Travel and Transport” category dominates, regardless of whether it is a workday or a weekend. To recognize is further that there are a large number of accommodations gathered near Kyoto Station which might also generate tweets in the travel category. **Figure 8** also shows that “travel and transport” is the most important type of activity in the Kyoto Station area regardless of the seasonal period. Its share has been maintained at around 70%, much larger than other types. It is interesting to note that the second most frequent tweet category was “Retail” during the State of Emergency, while “Dining and Drinking” were more frequent tweet locations in other periods.

We next consider another downtown area. The Kawaramachi area is known for dining, entertainment, and shopping facilities. According to **Figure 9** [B], it can be seen that “Dining and Drinking” in the Kawaramachi area has a low difference between lunch and dinner peaks from 15-18 on weekends. Similar to our discussion for **Figure 5**, this is an indication for an area with cafés and extended mealtimes on weekends. **Figure 10** shows that, except for the State of Emergency, that “Dining and Drinking” tweets account for almost half of the activities, while it only accounts for 36% during the “State of Emergency”, where “Retail” takes the largest share with 40%. This indicates that the gastronomical industry in the Kawaramachi area is highly affected by the State of Emergency declaration policy.

Finally, we investigate tweets in the Gion area for which the Foursquare POIs were already shown in **Figure 2**. The area is in general known for some sights such as “Yasaka Jinja” as well as high-end customers' nightlife activities.

This explains why we find in **Figure 11**, comparing [A] and [B], that “Dining and Drinking” as well as “Landmarks and Outdoors” in the Gion Area are prolonged on weekends compared to workdays, surpassing “Travel and Transport” as the dominant type on workdays. Moreover, the 24h distribution of “Dining and Drinking” and “Landmarks and Outdoors” in this area is positively correlated, with “Dining and Drinking” having only a lunchtime peak on weekends. This suggests that in particular at weekends some demand for dining in this area is brought by visits to landmarks, while, in general, more people move to other areas for dinner at night, such as the fairly nearby Kawaramachi area, where there is a wider range of restaurants.

Finally, as shown in **Figure 12**, the tweet share of “Landmarks and Outdoors” as well as “Community and Government” is greatly reduced during the State of Emergency. While this can be expected, in connection with our above discussion, we suggest this can again serve as an index for the economic damage caused by this policy.

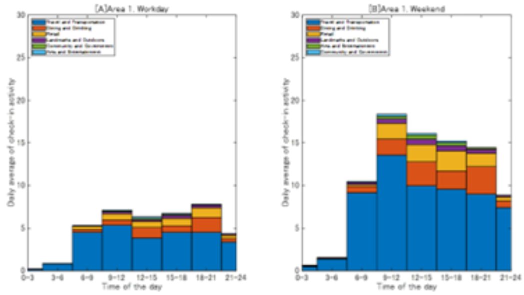
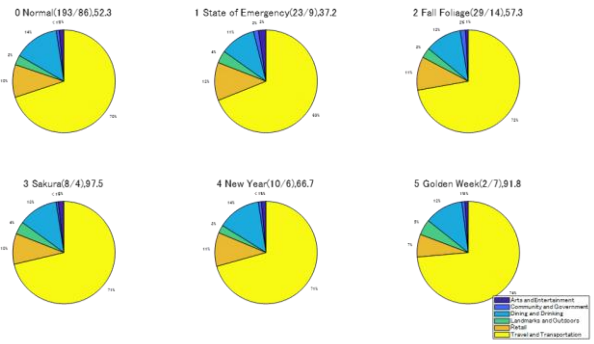


Fig.7 24h distribution of check-ins in the Kyoto Station area.



* The number behind the name of the season indicates the Daily Average Number of Tweets.

Fig.8 Activity type share of every seasonal period in the Kyoto Station area.

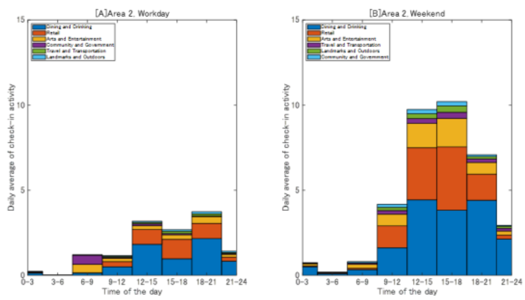


Fig.9 24h distribution of check-ins in the Kawaramachi area.

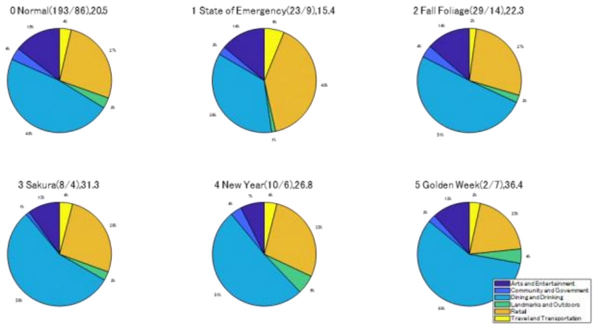


Fig.10 24h distribution of check-ins in the Kawaramachi area.

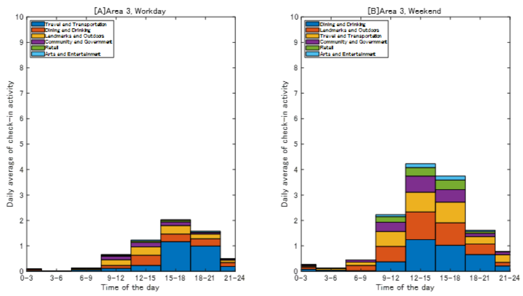


Fig.11 24h distribution of check-ins in the Gion area.

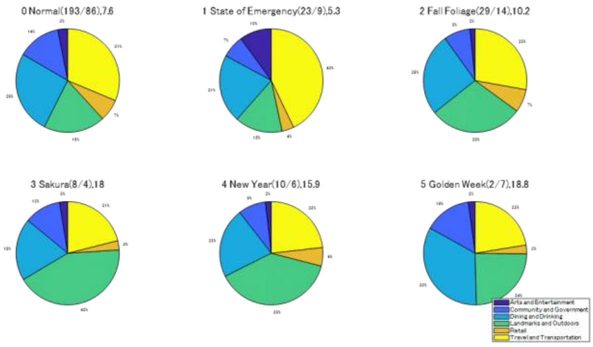


Fig.12 Activity type share of every period in the Gion area.

6. DISCUSSION AND CONCLUSION

We are aiming to infer the activity type of tweets to understand the current status of people's activities and the purpose of visiting different regions. We found that the check-ins shared from Swarm occupy a large proportion of the geo-tagged tweets and can be used to infer the type of activity participation. After cleaning the meaningless duplicate posts, we matched the check-ins with POIs and analyzed the percentage of activities on the Foursquare activity “parent category” level. We obtained six categories of activities providing us with detailed activity distributions. We can conclude that it is feasible to analyze the activity share by the time of day as well for seasonal differences at the city level, as well as at a more local level for some areas where the tweet number is sufficient. Clearly, the tweet sample is biased towards young, outgoing persons and one must be careful when aiming to infer directly from the tweets the actual amount of activities that have been conducted. What we propose is that the biases in the data set remain stable over the data collection period. If that is true, then the tweets allow us to capture trends. In other words, the relative differences between areas, times of day, and seasons are more meaningful. We note, however, that additional analysis, omitted for brevity, suggests that the tweet distribution of some popular specific locations also is in line with the “Google Popular times” busyness bar of those places.

Our analysis shows in all areas throughout Kyoto city a decrease in the total number of activities under the fourth Declaration of State of Emergency, a measure of the Japanese government in response to the COVID pandemic. Accordingly, the social activities inferred from tweets at dining, entertainment, and recreational places are particularly reduced whereas retail and transport showed a smaller decrease. We could further identify correlations in activities indicating that persons engage in multiple activities in the same area. If one breaks the analysis further down to particular POIs, this can help planners to understand the touristic and economic value of particular places for the whole area. As part of this, our analysis could further indicate the growth in visitors and particular activities during the peak touristic seasons. Our results provide one indication as to how much activities have recovered in the latter stages of the COVID pandemic. With even longer-term data we suggest the Twitter data can hence contribute to understanding the “new normal” in terms of changed activity distribution.

Our analysis also highlighted the limitations of the Twitter data. Aiming to analyze an activity distribution on a too fine-grained temporal and spatial scale leads to few data for meaningful analysis. Further, generally we suggest that the longer the time over which the data collection is continued, the more the analysis can gain in value. Another specific further work direction for us is that we consider exploring a regrouping of the locations shown in **Table 2** into different parent categories than those given by Foursquare. In a Japanese context, it appears more meaningful to regroup the categories “Community and Government” and “Landmarks and Outdoor”. For example “Temple” and “Shrine” might be better grouped together with “Historic and protected site”. We also consider analyzing the activity chains of some active users and note that our analysis could be extended beyond Kyoto. The fact that the data are, at least currently, freely available and that Twitter is used all over the world makes it also feasible to repeat our analysis in other countries.

ACKNOWLEDGMENT: This work was supported by the EIG CONCERT-Japan DARUMA project, Grant Number JPMJSC20C4 funded by JST SICORP, Japan. This work was also supported by Project PCI2020-120706-2 (DARUMA), funded by MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR, within the international research program EIG-CONCERT-JAPAN

REFERENCES

Aguilar, M. (2014). “Foursquare's Swarm App Tells Your Friends Generally Where You Are”. GIZMODO. <https://gizmodo.com/foursquares-forthcoming-swarm-app-ditches-check-ins-for-1570286006>. (accessed 2023-01-09)

FPO (2023). “Categories”. Foursquare documentation. Available from <https://location.foursquare.com/places/docs/categories>. (Accessed 2023-02-01).

García-Palomares, J. C., Salas-Olmedo, M. H., Moya-Gómez, B., Condeço-Melhorado, A., Gutiérrez, J. (2017) City dynamics through Twitter: Relationships between land use and spatiotemporal demographics Cities.

Kobayashi, H. (2022) On the relationship between crowdsourced sentiments and mobility trends during COVID-19 : A case study of Kyoto. Kyoto University, Faculty of Engineering, Graduation Thesis. (in. Japanese)

Lee, R., Wakamiya, S., & Sumiya, K. (2012) Urban Characteristics Extraction Based on Crowd Behavior by Tweets Analysis. Journal of Information Processing Society of Japan: database (TOD), Vol. 5, No. 2, pp. 36–52, 2012. (in Japanese)

Martín, Y., Cutter, S.L., Li, Z., Emrich, C.T., Mitchell, J. T. (2020) “Using Geotagged Tweets to Track Population Movements to and from Puerto Rico After Hurricane Maria.” Population and Environment.

Noulas, A., Scellato, S., Mascolo, C., & Pontil, M. (2011) An empirical study of geographic user activity patterns in foursquare. In ICWSM '11: International Conference on Weblogs and Social Media

Pabilonia, S. W., & Vernon, V. (2022). Telework, wages, and time use in the United States. Review of Economics of the Household, 20(3), 687-734.

Shen, K., Schmöcker, J. D., Sun, W., & Qureshi, A. G. (2022). Calibration of sightseeing tour choices considering multiple decision criteria with diminishing reward. Transportation, 1-25.

Twitter, Inc. (2023) “Tweet location FAQs”. Twitter Help Center. <https://help.twitter.com/en/safety-and-security/tweet-location-settings>. (accessed 2023-01-16)

(Received March 3, 2022)