Mining heatstroke hotspot for inbound tourists in Tokyo

Tianqi Xia, Satoshi Miyazawa, Xuan Song, Yuki Akiyama, Renhe Jiang, Kyoung-sook Kim and Ryosuke Shibasaki

Abstract: Summer is the popular season for inbound tourism in Tokyo during which a lot of events will be held. Nevertheless, as the city with four distinct seasons, the tourists who visited Tokyo during summer vocation are likely to suffer from heatstroke when enjoying outdoor traveling. In order to help local governments improve travel quality and better prepare for the upcoming 2020 Tokyo Olympic Games, this research aims to extract tourism regions that have a higher risk of heatstroke via heterogeneous data. First, the candidate POIs are extracted from multiple data sources, then the popular ROIs are generated by matching POI datasets to several ROI datasets. With the regions generated by routes and POIs, the risk of heatstroke is evaluated by several indicators.

Keywords: Heatstroke; Inbound visitors; Data mining; Spatial analysis

1. Introduction

With the policy of developing a tourism country and the upcoming 2020 Tokyo Olympic Games, Tokyo, as the capital city of Japan with abundant tourism resources and events, attracts more and more inbound tourists during the recent years. At the same time, however, as a metropolis that has four distinct seasons and suffers from the urban heat island effect, Tokyo has a high risk of heatstroke during summertime, which is the peak season for inbound tourism and is exactly the time that Olympic Games will be held.

Heatstroke is one kind of life-threatening emergency caused by prolonged exposure to high temperature. The previous studies of heatstroke mainly analyzed heatstroke risk at a micro-scale with different indicators from the perspective of hazard, vulnerability, and exposure. (Kasai, 2017) Particularly, exposure refers to the number of people or the amount of time exposed to the high temperature. In the previous studies, exposure is always represented by the census data of local residents. Nevertheless, inbound tourists, though having a long time of outdoor activities that indicating a longer temperature exposure time than the residents, are not included by the census, which makes it difficult to analyze their heatstroke risk. Thus in order to provide better tourism service for inbound visitors and better prepare for the Tokyo Olympic Games, it is necessary to find out those potential regions that are popular among foreign visitors, but at the same time require long time outdoor mobility.

With the concerns mentioned above, the objective of this research is to extract tourism regions that have a higher risk of heatstroke via heterogeneous data. First, the candidate tourist attractions are extracted based on their popularity and types. Then the extracted POIs are matched to ROIs from different data sources. Finally, with the generated ROI, we utilize several indicators to evaluate their heatstroke risk and extract the hotspots based on the indicators.

2. Data and Processing

Data collection and processing in this research are challenging due to the fact there is no perfect dataset for representing the ROI of inbound visitors. Thus in this study, we firstly review several kinds of datasets that we used in this research and then introducing the

Tianqi Xia, National Institute of Advanced Industrial Science and Technology.
xiatianqi@csis.u-tokyo.ac.jp

processing approach for generating ROIs based on the POI dataset as well as the extra dataset and process utilized for measuring the indicators.

2.1 Point data and POI

Point data is the common spatial representation of inbound tourism data. Generally, check-in point data and the tourist attraction data from travel service website can represent the POIs as well as their popularity.

In this study, we use both types of point data to extract POI information. For SNS data, we utilize the tweet data of inbound visitors collected by Nightley Inc. during the summertime in 2018. For representing a more precise foreign visitor distribution during high temperature, we match twitter data to the weather information in 2018 based on the timestamp and extract those tweets that are tweeted over 28 °C with more than 0 solar radiation. The number of filtered tweets is 5923 in Tokyo 23 wards.

In addition, we also collect tourism attraction information from TripAdvisor including its comment numbers in foreign languages as well as its spatial information represented by address and coordinates. In Tokyo 23 wards, there are around 19687 tourist attractions which including over 15 categories including parks, historical relics, entertainment facilities and restaurants. To simplify the preprocessing and extract those more meaningful ROIs, we only utilize the POIs that contain more than 5 comments in foreign languages.

2.2 ROI data

An ROI represents a region that can be potentially visited by the inbound tourists. Generally, outdoor ROIs could be either with a distinct boundary, such as campuses and parks or with a vague boundary such as the commercial area. For the former one, the ROI usually refers to one main POI which can represent its boundary and sometimes several POIs that belongs to the main POI (e.g. Ueno Park and Ueno Toshogu), while for the latter, the ROI always consists of several independent POIs that are concentrated in some specific area. Thus in this study, for POIs which are correspondent to ROIs with specific boundary, we match them to the polygon data in OpenStreetMap; while for the ROI with vague boundaries which are mainly the entertainment and shopping clusters, we match them to the commercial accumulations adapted from the work of Akiyama (2018) with the data provided by JoRAS.

2.3 Route data

In this study, for the outdoor activity of inbound tourism, we consider about three types of walking routes including routes from railway stations to the ROIs, the route within each ROI and the routes between ROIs. All walking routes are acquired from OpenStreetMap data with the tool OSMnx developed by Boeing (2017). The routes between station and ROI and between ROIs refer to the shortest route from the origin to the destination while for the inner ROI route, we represent it by the maximum distance of the shortest path between any nodes within the ROI. Thus, for those ROIs with no paths, we extract the minimum bounding rectangle (MBR) of the ROI and choose the longer edge as the potential maximum route length of the ROI.

2.4 Data matching

Data matching and fusion are fundamental in this research to combine the spatial range with the popularity of each ROI for evaluating the heatstroke risk. For those ROI with distinct boundaries in OSM, we match POI groups to ROI dataset by spatial analysis, address matching and name pattern matching.

On the other hand, matching POI data to commercial accumulations mainly based on the spatial information in these two datasets. Since the ROI calculated in JoRAS include some daily life services, we choose the commercial facilities that only within a threshold distance of the POI and exclude the facilities that belong to daily life services. Furtherly, we utilize the DBSCAN method based on both buffer polygon and road network to merging the accumulations into larger commercial clusters.

For both kinds of ROI, in the current research, we extract those regions with an area over 1 hectare to remove those ROIs that are potentially indoor shopping malls or indoor theme parks. Totally, we extract 111 outdoor ROIs from OpenStreetMap and generate 112 commercial accumulations. The spatial distribution of these two ROIs is visualized in Figure 1. From the figure, we can find that in the downtown of Tokyo, the ROIs are more concentrated than in the peripheral area. In the downtown of Tokyo tends to generate larger commercial accumulations due to a dense distribution of commercial facilities.

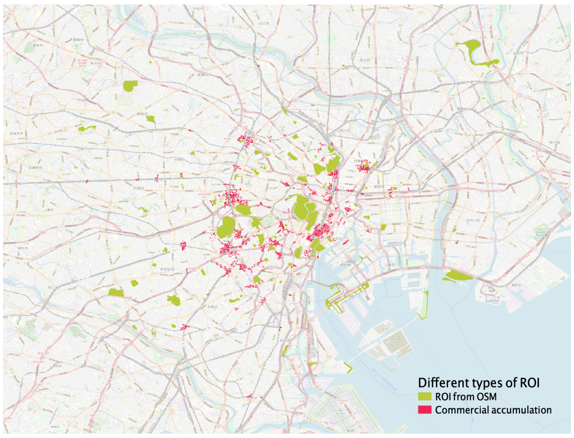

Figure.1 Spatial distribution of different types of ROI

3. Heatstroke exposure risk analysis

Since exposure risk concerning both the volume and time of outdoor activity, therefore in this study, we look into both the potential exposure tourist numbers and potential exposure time to analyze heatstroke exposure risk.

The number of potential exposure tourists is measured by two indicators: the number of comments in each ROI represents the popularity of the ROI by summing up the whole foreign comment numbers of all POIs belong to the ROI. The number of high-temperature tweets represents the observational result of inbound visitor numbers. On the other hand, the exposure time for the tourists should be proportional to the traveling distance if assuming tourists travel at an average speed. As is mentioned before, in this study we represent the traveling distance by the route from the nearest station to an ROI, the route distance within an ROI and the distance from an ROI to the neighboring ROIs.

The result of each indicator is visualized in Figure 2. From the figures, we can find that these indicators show different spatial distribution features for us to find the hotspot under each indicator: comparing ROI comments with tweets, we can find that though their overall patterns are similar, the ratio of tweets in parks get smaller while is better to reflect the potential move around big shopping malls such as the ROI in Odaiba. For inner travel route distance, the result reveals that in a lot of parks the potential time cost would be large for a deep visit, which indicate that it is necessary to plan for a detailed travel route that is less risky in high temperature. Railway station distance and neighboring ROI distance show an opposite distribution pattern for that central Tokyo has a large density of railway stations and ROIs, thus though the visitors can spend less time on moving from station to the destination, they have more potential possibility to walk to the neighboring ROIs for another visit, which also expand the exposure time. Finally, the result for summing up the logarithm of each indicator shows that the ROIs in central Tokyo has a higher comprehensive exposure risk.

4. Conclusion

In this study, we extracted tourism ROI that have a higher exposure risk of heatstroke utilizing heterogeneous data and utilize several indicators to evaluate those regions. The contribution of this study

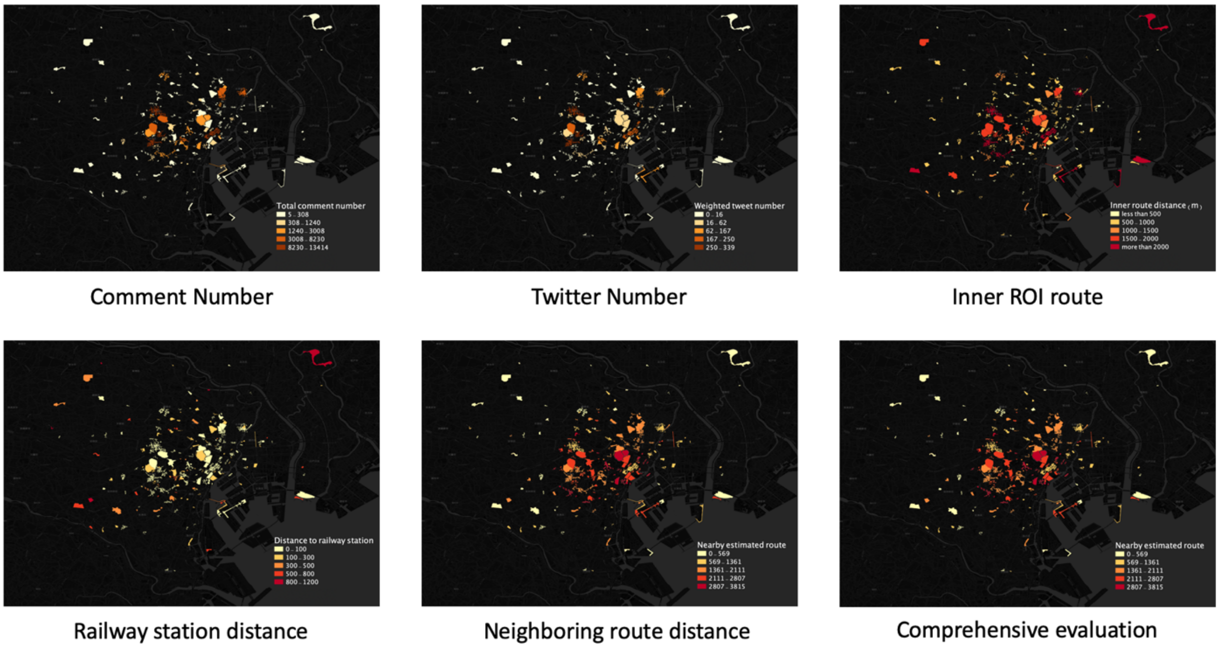

Figure.2 Visualization of indicator spatial analysis result

is that it focuses on the research target that is important while usually neglected in other studies, and this study builds a whole system from data collection, matching, analysis and visualization.

The limitation of this research is that the indicator is still not enough for analyzing heatstroke. In the current stage, we do not consider the detail of land type in each ROI, which might also influence the risk from other perspectives like hazard and vulnerability. In addition, the performance of ROI extraction would be difficult to be evaluated. This is especially a problem for the downtown of Tokyo with high POI density.

In the future, we will consider modifying the commercial area extraction part via utilizing the street data and hopefully the image data to further analyze heatstroke risk and walkability.

Acknowledgment

This research project was based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

Kasai, M., Okaze, T., Yamamoto, M., Mochida, A. and Hanaoka, K., 2017. Summer heatstroke risk prediction for Tokyo in the 2030s based on mesoscale simulations by WRF. *Journal of Heat Island Institute International* Vol, 12, p.2.

Yuki, A., 2018. Monitoring of the Spatial Distribution of Commercial Accumulations throughout Japan Using Micro Geodata. *E-journal GEO*, 13(1), pp.109-126.

Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, pp.126-139.

Sekimoto, Y., Shibasaki, R., Kanasugi, H., Usui, T. and Shimazaki, Y., 2011. Pflow: Reconstructing people flow recycling large-scale social survey data. *IEEE Pervasive Computing*, 10(4), pp.27-35.