Estimating Evacuation Hotspots using GPS data: What happened after the large earthquakes in Kumamoto, Japan?

Takahiro Yabe
University of Tokyo[1]
yabe0505@iis.u-tokyo.ac.jp

Kota Tsubouchi
Yahoo! Japan Research[2]
ktsubouc@yahoo-corp.co.jp

Akihito Sudo
University of Tokyo[1]
sudoa@iis.u-tokyo.ac.jp

Yoshihide Sekimoto
University of Tokyo[1]
sekimoto@iis.u-tokyo.ac.jp

ABSTRACT
Kumamoto prefecture, Japan, was hit by enormous (Magnitude 6.5, 7.3) earthquakes on 14th and 16th of April, 2016. As a result of the shocks, more than 10,000 buildings were severely damaged, and over 100,000 people had to evacuate away from their homes. After the earthquake, it took the authorities several days to grasp the locations where people were evacuating, which delayed the distribution of supply and rescue teams. This situation was made even harder since some people evacuated to places that were not designated as evacuation shelters. Conventional methods for grasping evacuation hotspots require field surveys, which take time and are also difficult to execute right after the hazard in the confusion. We propose a framework to efficiently estimate the evacuation hotspots using location data collected from mobile phones. To validate our framework, we estimated the locations that were congested with evacuees after the Kumamoto earthquake using GPS data collected by Yahoo! Japan. We also verified that our estimation results were very high, by checking the features located in each grid with high anomaly value. Moreover, for one of the non-designated evacuation hotspots, we accurately estimated the population transition of before and after the earthquake, which we validated using newspaper reports.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Keywords
Disaster Management, Human Mobility, Urban Computing, Evacuation Activities, GPS Data

1. INTRODUCTION
At 9:26PM of 14th April 2016, Kumamoto prefecture, located on Kyushu Island of Japan, was struck by a M6.5 earthquake. This shock was followed by another M7.3 earthquake next day, at 1:25AM, 16th April 2016. Since the first shock, over 1200 earthquakes larger than seismic level 1 has occurred in this area [1]. As a result of the two large earthquakes and many minor shocks, more than 10,000 residential buildings had collapsed, and unfortunately 49 people were killed mainly due to building collapse. As we can observe from Figure 1, the second and largest shock struck near the densely populated central areas of Kumamoto. This forced over 100,000 people to move away from their homes to evacuation facilities.

The mass evacuation activities of the victims caused serious issues. Grasping the locations of all the evacuation hotspots was extremely difficult in the chaos and confusion after the earthquake. This was made even harder since some people evacuated to locations which were not officially designated as evacuation places, such as parking areas of large shopping malls. As a result, many evacuation hotspots which were not recognized by the administrative organizations as evacuation shelters couldn't be provided with food and supplies efficiently [2]. This increased the burden for the evacuees for several days. There is an urgent need for an efficient framework for estimating evacuation hotspots right after a natural disaster. The framework needs to require less time and less work load for the authority members who are busy managing the situation after the earthquake, compared to the conventional on-foot search for evacuation hotspots.

Recently, GPS and call detail records (CDR) of mobile phones are being used for human mobility analysis [3,4,5,7], and is applied to various fields of study such as traffic management [11,12,14],

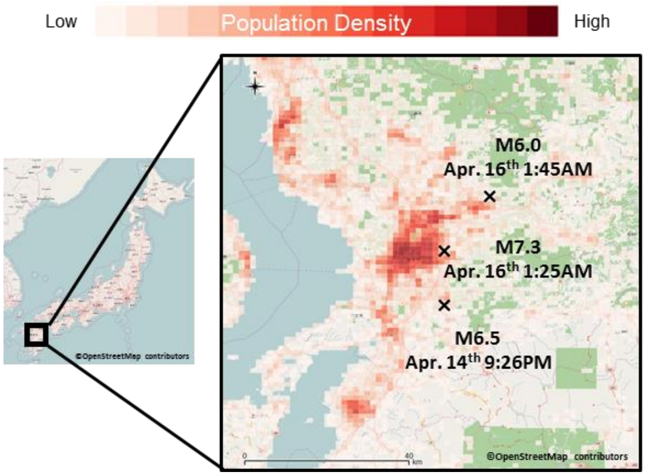

Figure 1. Crosses indicate epicenters of the three >M6 earthquakes. The thickness of red color in each grid represent the usual population density of each grid. The largest earthquake occurred near the highly populated area.

[1] 4-6-1 Komaba, Meguro-ku, Tokyo, Japan. (81)-3-5452-6009

[2] 9-7-1 Akasaka, Minato-ku, Tokyo, Japan. (81)-3-6864-3412

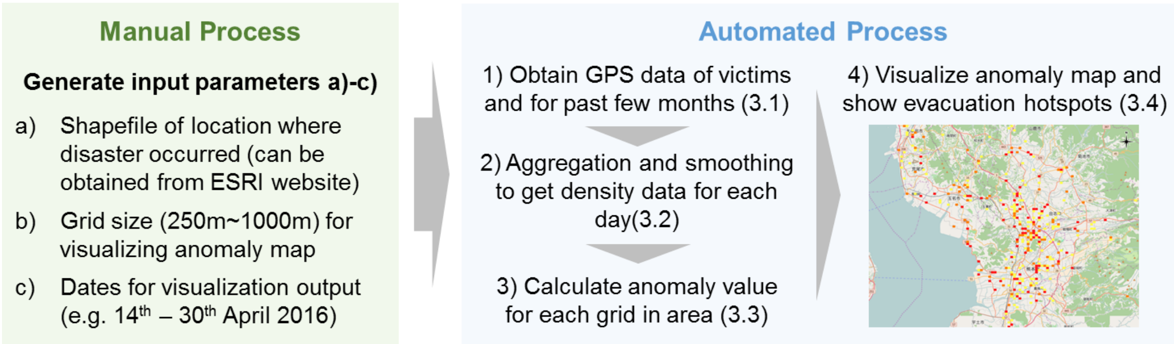

Figure 2. Diagram of our proposed framework. After the occurrence of the disaster, we manually input the a) shapefile of location where disaster occurred, b) the preferred grid size for visualization, and c) the dates for the visualization output. All the following processes are automated. The system calculates the anomaly value for each grid and visualizes the evacuation hotspots. This output contributes to a more efficient supply and rescue distribution plan. The numbers in brackets are section numbers explaining that part of the framework.

urban planning [13], and pandemic simulations [16]. Some studies have analyzed the irregular human mobility after natural disasters such as Hurricane Sandy, Great East Japan Earthquake, and Haiti Earthquake [17,18,19], but none have proposed a method for a real-time evacuation hotspot estimation.

In this paper, we propose a framework for estimating evacuation hotspots using mobile phone GPS data. We used Yahoo! Japan's GPS dataset for analyzing the evacuation hotspots after the Kumamoto earthquake. Through the case study in the Kumamoto earthquake, we show our framework can accurately estimate evacuation hotspots, and also that this process can be completed at a significantly high speed and low effort compared to the conventional on-foot investigations.

Our key contributions of this paper are as follows:

- We propose a framework to estimate evacuation hotspots following large natural disasters by using mobile phone GPS data.
- We validate our framework by estimating the evacuation hotspots after the Kumamoto earthquake using actual GPS data.
- We verify that the estimations are precise by comparing the population transition in one of the evacuation facilities, to the information obtained through newspapers and reports.

2. Related Works

2.1 Human Mobility Analysis

The increasing availability and fusion techniques of big and heterogeneous data has enabled the analysis for tackling major issues that cities face [8,9]. Due to the popularization of mobile phones, call detail record (CDR) and GPS data have become popular for analyzing people movement [3,4,5,7] and nation-wide population distribution [4]. Other studies have applied human mobility analysis to many fields such as traffic estimation [11,12,14], urban planning [13], and public health [16].

2.2 Disaster Mobility Analysis

Recently, to understand the irregular people movement under disaster conditions, mobile phone CDR/GPS data are used for analyzing calling activities after large disasters [6]. Human evacuation activities after natural disasters such as the Haiti earthquake [17], Hurricane Sandy [18], and the Great East Japan Earthquake [19] are also analyzed. However, these studies are commonly the analysis of people movement after a disaster and do not attempt to infer the real-time evacuation hotspots after the disaster.

2.3 Frameworks for Estimating Evacuation Hotspots

Chen *et al.* [10] proposed a framework with agent based simulation for predicting human mobility after disasters using CDR, however used a simple potential model for the mobility prediction and did not attempt to estimate the evacuation hotspots. Horanont *et al.* [21] stated the potential usage of GPS data in emergency situations, but did not propose a method for evacuation hotspot estimation. We also have to note that the analysis was done after the settlement of the disaster in these two studies, and their frameworks' ability for prompt analysis and visualization have not been verified.

3. Proposed Framework

As shown in Figure 2, our proposed framework is consisted of the manual parameter input process and the automated process. The manual parameter input could be completed momentarily, since shapefiles are available online. The automated process is consisted of 4 parts; 1) location data collection, 2) aggregation and smoothing the GPS data, 3) calculating the anomaly value of each grid, and 4) visualization of estimated evacuation hotspots. This framework is performed iteratively after the occurrence of the disaster, to update the locations of evacuation hotspots. Our framework is efficiently designed so that it could be quickly processed, and also has little burden for the framework users since most of the parts are automated.

3.1 Location Data Collection

GPS data of the Yahoo! Japan app users are collected continuously, and they are stored in an internal server. Once a disaster occurs, the logs recorded within a period of a few months before the disaster are collected. Then, logs located near the disaster hit area are extracted for each day. For spatial extraction, shapefile data of the area of disaster occurrence are needed as input data. This data collection phase can be completed in a few hours.

3.2 Aggregation and Smoothing of GPS Data

We aggregate the collected GPS data to obtain the population distribution of each day. To calculate the population distribution at night, we aggregate the last log for each ID (where he/she is sleeping) into 500m~1000m grids to maintain the privacy of the users. Then, to overcome the relatively low sample rate (1.2%), we perform a kernel density estimation [22] given by equation (1).

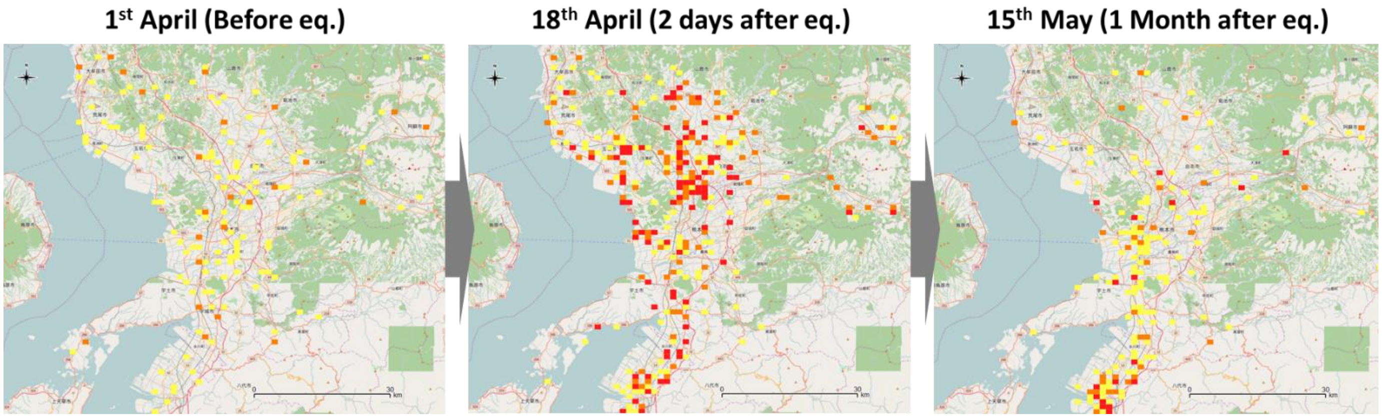

Figure 3. Map of Kumamoto area; before, 2 days after, and 1 month after the earthquake. The grids with anomal values K>3, 2<K<3, and 1<K<2 are colored red, orange, and yellow, respectively. We can observe the increase in irregularly congested grids right after the earthquake, and its decrease as time passes and start to get back to normal from the shock.

$$f_h(x) = \frac{1}{n * h} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-x_i}{h})^2} \qquad (1)$$

Where there are n grids in the Kumamoto area, x is the original population value, h is the bandwidth, and $f_h(x)$ is the estimated population value. We aggregate and smooth the population distribution data for all the days where data are available.

3.3 Calculation of Anomaly Value

The anomaly value $K_{i,j}$ of each grid j on day i in the Kumamoto area are calculated by equation (2), using the average population μ_j, the standard deviation σ_j for each grid j, and the population in grid j on day i, denoted by $M_{i,j}$. This anomaly value measures the relative congestion of the grid compared to an average day, also considering the fluctuation of the population in that grid. Therefore, the anomaly value $K_{i,j}$ is dependent on the divergence of the grid's usual population.

$$K_{i,j} = \frac{M_{i,j} - \mu_j}{\sigma_j} \qquad (2)$$

3.4 Visualization of Evacuation Hotspots

After calculating the anomaly value of the grids in the disaster-hit area, we visualize the anomaly grids onto a map, so that policy makers and administration organizations can check the evacuation hotspots easily. Visualization will be performed on a free GIS software called QGIS[3], and OpenStreetMap[4] would be used as the background map.

4. Experiment on Kumamoto Earthquake

4.1 GPS Dataset

The Yahoo! Japan Disaster App collects the GPS data of each individual, who have agreed to provide their location data to Yahoo! Japan when installing the app. Each GPS record contains an anonymized user ID, longitude, latitude, and timestamp. The GPS data are collected every day, when the smartphone is turned on, and when the individuals move around. In total, GPS data of around 1 million individuals (sample rate around 1% from all over Japan) have been collected, which makes Yahoo! Japan's GPS dataset one of the richest datasets in the world. As shown in Table 1, for the experiment, we used a total of 22,124 users' 418,119 total GPS logs from a period of January 1st to May 16th of 2016 which were located in Kumamoto area.

4.2 Visualization of Anomaly Map

Figure 3 shows the map of Kumamoto area with grids with K>3 colored in red, and 2<K<3 colored in orange. We can observe very few anomaly grids on the 1st April before the earthquake, meaning that the majority of the grids have a population within usual range. However on the 18th, after the large earthquakes, we can observe a significant increase in the number of anomaly grids, especially near the city center and the southern part when many people evacuated. These grids indicate the "evacuation hotspots", where people evacuated at a significant rate compared to the usual population in that grid. It is also interesting to observe K>3 grids located on roads passing near the coastline. The high anomaly values in these grids infer that many people stayed in their cars away from their houses. However, after a month from the earthquake, we can see a decrease in congested areas in Kumamoto area. This implies that many evacuees returned home (if their house was not completely damaged) or moved away to other areas of Japan for shelter.

Figure 4 shows the temporal transition of the percentage of K>3 and K>2 grids. By analyzing the anomaly value of usual days, we found that the probability distribution of usual anomaly values follow a normal distribution. Therefore, before the earthquake, around 0.2% of the grids have K>3, and around 3% of the grids have 2<K<3. However after the earthquake, more than 7% of the grids had a K>3, and 8% had 2<K<3, resulting in a significant increase in high density evacuation hotspots compared to usual days. We can clearly see the impact of the earthquakes, and also how the situation gradually returns to normal state as time passes.

Table 1. Number of unique IDs, number of logs of GPS data

Period of data	Average daily number of IDs in Kumamoto area	Average daily total GPS logs in Kumamoto area
2016/01/01 ~ 2016/05/16	22,124 (1.2% sample rate)	418,119 (avg. 19 logs/user/day)

[3] http://www.qgis.org/en/site/

[4] https://www.openstreetmap.org/

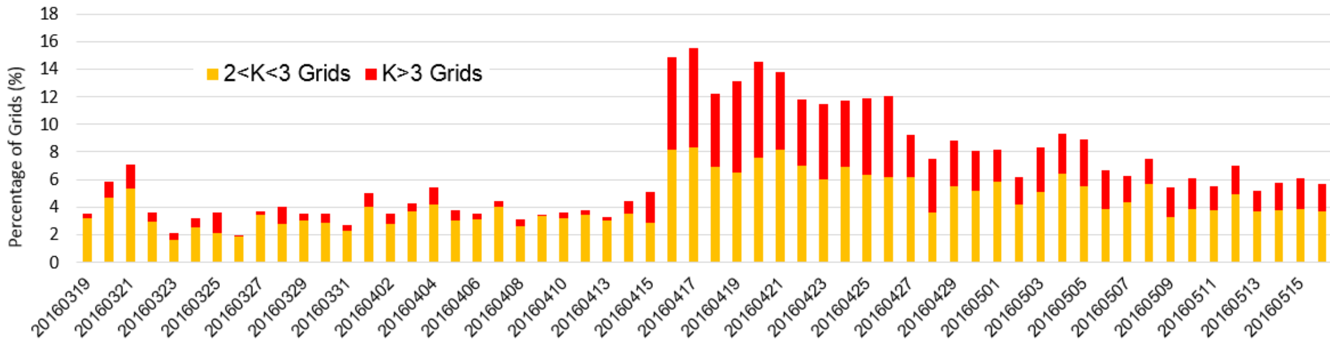

Figure 4. Percentage of K>3, 2<K<3 anomaly grids by days. In normal days before the earthquakes, around 1% of K>3 and 3% of 2<K<3 grids are observed. However, after the shocks, irregularly congested grids increase at a significant rate. As time passes, the number of anomaly grids decrease, and by May 15th, the percentage of anomaly grids almost reach normal state.

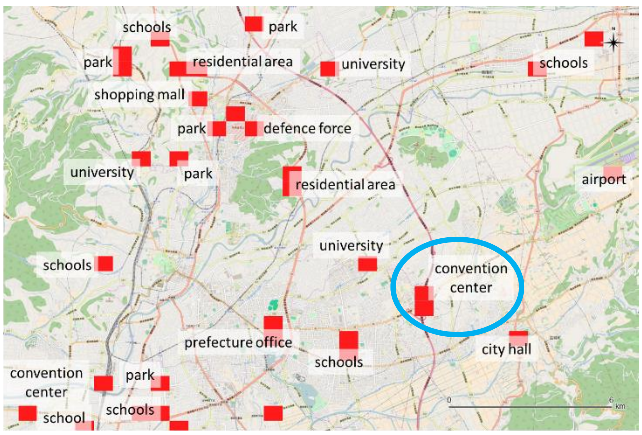

Figure 5. Map of central Kumamoto with K>3 grids colored in red. Close investigation on these grids revealed that most of these grids contain features that have a large capacity for evacuees such as schools, city halls, and convention centers. The blue circle shows Grand Messe Kumamoto (GMK).

Using our method, not only were we able to estimate the evacuation hotspots after the earthquake, but also the impact of the earthquake and the length of time it takes for a city to recover from a certain type of disaster.

4.3 Validation of Estimated Hotspots

Figure 5 shows a map of central Kumamoto and the grids with K>3 colored in red. We validated our estimation of evacuation hotspots by checking the features located in each grid. The type of feature located in each of the grids are written beside the grid. Many of the estimated hotspots contained features that have the capacity to contain large evacuation population, such as schools, city halls, convention centers and parks. From these results, we can conclude that our framework successfully estimated the evacuation hotspots after the earthquake. To verify our results more, in the next subsection, we will focus on a convention center called Grand Messe Kumamoto, which was not designated as an evacuation shelter before the earthquake but hosted a large number of evacuees, and analyze the transition of population density in that particular grid.

4.4 Case Study of Grand Messe Kumamoto

To analyze the evacuation activities with more detail, we focus on one evacuation hotspot where K>3 anomaly was detected, and plot the transition of the daily population in that selected hotspot. We

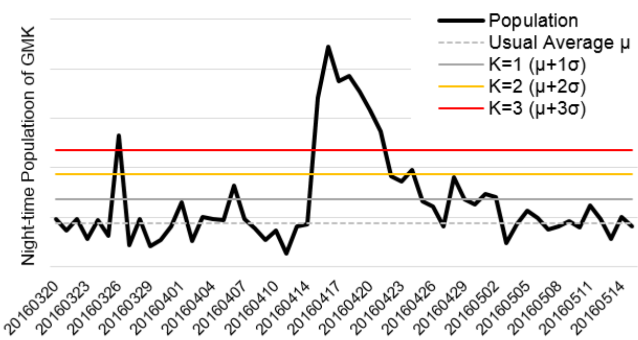

Figure 6. Transition of population density near "Grand Messe Kumamoto", a large shopping mall which was not designated as an evacuation place. We can observe high congestion of K>3 after the earthquake, on April 15th ~ 21st.

focused on "Grand Messe Kumamoto (GMK)" (circled in blue in Figure 5), a convention center in Kumamoto area, and analyzed its daily transition of population in that facility. According to newspaper articles [20], many evacuees gathered in the parking area of GMK right after the earthquake, despite the fact that this facility was not designated as an evacuation location.

The broad black line in Figure 6 shows the daily transition of the night-time population in GMK from March 20th to 14th May 2016. The dotted gray line denotes the average population μ of usual days in GMK, and the gray, orange, and red lines indicate the K=1, K=2, K=3 lines for GMK, respectively. We can observe a rapid increase after the earthquakes on April 16th, and a significant anomaly in population density for more than a week over K=3. We can also spot an instantaneous K>3 on Saturday, March 26th. There was actually a large music festival on this day at GMK, which is an example of an anomaly within usual days. After April 18th, we can observe a decrease of population in GMK, and a gradual return to a usual level of population. By the beginning of May, the population in GMK has transferred back to the normal state.

The increase of population on April 16th coincides with the information on the newspaper article [20]. We can conclude that the population in GMK, one of the evacuation hotspots, was accurately inferred.

5. Discussion

Our framework for estimating evacuation hotspots using GPS data can provide useful information to administrative organizations quicker and with less effort than conventional methods. Providing useful and accurate information can contribute to making efficient supply distribution and rescue operation plans after disasters.

By calculating the anomaly value of each grid in Kumamoto after the earthquake, we were able to estimate the distribution of irregularly congested grids. Also, we were able to observe the significant increase in number of grids with high anomaly values right after the occurrence of the earthquake. Also, the gradual decrease of grids with anomaly values showed the settling down of irregularity after a few weeks from the occurrence of the earthquakes. Through the analysis, we were able to observe the impact of earthquakes on the people's evacuation activities.

Checking the features in each grid verified the accuracy of our framework. We were able to even detect locations where administrative organizations had not designated as evacuation shelters, such as Grand Messe Kumamoto.

Analyzing the daily transition of population in one of the non-designated evacuation hotspots has showed that our estimated population fluctuation pattern was similar to the newspaper report. Also, by comparing the peak with a usual event (music festival), we were able to relatively understand the level of congestion caused by the evacuation activities, which was significantly higher than the anomaly in usual state.

To decrease the processing time and burden for the people who use the framework, we automated most of the processes and only parameter input has to be done manually.

6. Conclusion

In this paper, we proposed a framework for estimating evacuation hotspots by checking each grid's anomaly value after large disasters, using mobile phone GPS data. To the best of our knowledge, this framework is the first to focus on estimating evacuation hotspots using GPS data. Our framework can function quicker and with less effort compared to conventional methods that involve on-foot searches for evacuation centers where people are gathering.

To validate our method, we analyzed the population density anomaly after the Kumamoto earthquake (M7.3), and observed the sharp increase of high anomaly value grids in Kumamoto area caused by the evacuation activities of the victims. We then verified our estimation by looking at the features included in each grid, and also newspaper articles that mentioned the population transition in one of the evacuation hotspots.

Through the validation case study of Kumamoto, we have confirmed the high accuracy of this framework's estimation. Also, it is quicker and requires low workload compared to conventional on-foot survey methods. We have constructed the system, and is planned to be actually used in the next disaster.

7. References

1. Japan Meteorological Agency, Evaluation of the Kumamoto Earthquake, (2016). (in Japanese)
2. West Japan Newspaper, Disparity between evacuation centers, April 20th (2016). (in Japanese)
3. Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. Understanding individual human mobility patterns. Nature, 453(7196), (2008).
4. Sevtsuk, A., & Ratti, C. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. Journal of Urban Technology, 17(1), 41-60. (2010).
5. Ashbrook, D., & Starner, T. Using GPS to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing, 7, 5, 275-286. (2003).
6. Bagrow, J. P., Wang, D., & Barabasi, A. L. Collective response of human populations to large-scale emergencies. PloS one, 6(3) (2011).
7. Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. Estimating origin-destination flows using mobile phone location data. IEEE Pervasive Computing, 10(4), (2011).
8. Zheng, Y., Capra, L., Wolfson, O., & Yang, H. Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), 38. (2014).
9. Zheng, Y. Methodologies for cross-domain data fusion: an overview. Big Data, IEEE Transactions on, 1(1), 16-34. (2015).
10. Chen, F., Zhai, Z., & Madey, G. Dynamic adaptive disaster simulation: developing a predictive model of emergency behavior using cell phone and GIS data. In Proceedings of the 2011 Workshop on Agent-Directed Simulation (pp. 5-12). Society for Computer Simulation International. (2011).
11. Demissie, M. G., de Almeida Correia, G. H., & Bento, C. Intelligent road traffic status detection system through cellular networks handover information, Transportation research part C: emerging technologies, 32, 76-88. (2013).
12. Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. Development of origin–destination matrices using mobile phone call data. Transportation Research Part C: Emerging Technologies, 40, 63-74. (2014).
13. Wang, P., Hunter, T., Bayen, A. M., Schechtner, K. & González, M. C. Understanding road usage patterns in urban areas. Sci. Rep. 2, 1001 (2012).
14. Yang, Y., Gerstle, D., Widhalm, P., Bauer, D. & González, M. The potential of low-frequency data for the monitoring and control of bus performance. Transport. Res. Rec. J. Transport. Res. (2013).
15. Pentland, A. Society's nervous system: Building effective government, energy, and public health systems. IEEE Computer 45, 31–38 (2012).
16. Colizza, V., Barrat, A., Barthelemy, M., Valleron, A. J., & Vespignani, A.. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med, 4(1) (2007).
17. Lu, X., Bengtsson, L., & Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. Proc. National Academy of Sciences, 109(29), 11576-11581. (2012).
18. Wang, Q., & Taylor, J. E. Quantifying human mobility perturbation and resilience in Hurricane Sandy. PLoS one, 9(11), (2014).
19. Song, X., Zhang, Q., Sekimoto, Y., Horanont, T., Ueyama, S., & Shibasaki, R. Modeling and probabilistic reasoning of population evacuation during large-scale disaster. Proc. ACM SIGKDD (2013).
20. Kumamoto Daily Newspaper, Evacuation center at Grand Messe Kumamoto, April 24th (2016). (in Japanese)
21. Horanont, T., Witayangkurn, A., Sekimoto, Y., & Shibasaki, R. Large-scale auto-GPS analysis for discerning behavior change during crisis. IEEE Intelligent Systems, (4), 26-34. (2013).
22. Tukey, J. W. Exploratory data analysis. (1977).