Paper:

Development of GIS Integrated Big Data Research Toolbox (BigGIS-RTX) for Mobile CDR Data Processing in Disasters Management

Ko Ko Lwin†, Yoshihide Sekimoto, and Wataru Takeuchi

Institute of Industrial Science, The University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
†Corresponding author, E-mail: kokolwin@iis.u-tokyo.ac.jp
[Received September 1, 2017; accepted March 1, 2018]

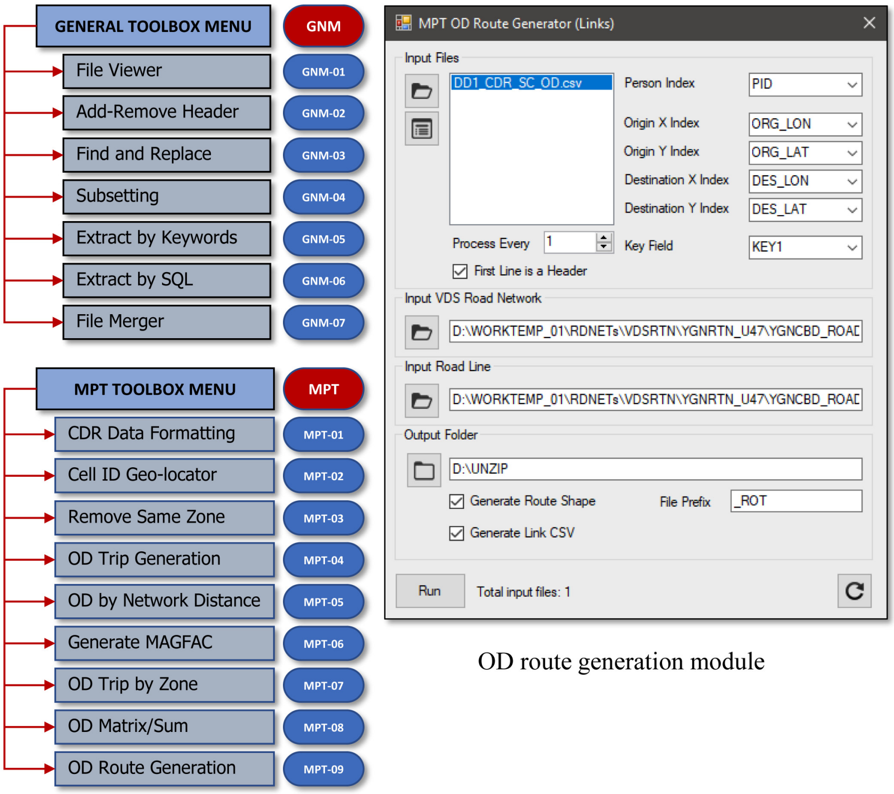
This article reports the development of a geographical information system (GIS) embedded text-based geospatial Big Data research toolbox (BigGIS-RTX) designed especially for mobile CDR (Call Details Record) data processing in urban transport planning and disaster management. BigGIS-RTX is a standalone computer program that aims to provide a bridge between geospatial Big Data and end users (i.e. students and researchers) by reducing difficulties in handling geospatial Big Data processing and analysis tasks. This research toolbox makes it possible to handle text-based geospatial Big Data cleaning, formatting, subsetting, and extraction by keywords or structured query language (SQL), CDR data aggregation by base transceiver stations (BTSs), generation of origin–destination (OD) trips, OD matrices, and OD routes, and computation of OD links. Moreover, this research toolbox can be integrated with current commercial GIS software to perform further geospatial analysis functions to improve spatial decision making in urban and transport planning and disaster management. In this report, we discuss two current research outputs using BigGIS-RTX: first, multitemporal grid square population estimation and second, human mobility studies in transportation planning. These research outputs are essential for disaster management and emergency preparedness in terms of providing knowledge and information about population distribution changes over space and time, human mobility flow by a user defined time frame, and travel volume or link count information for individual road segments. This research is part of the core project "Development of a Comprehensive Disaster Resilience System and Collaboration Platform in Myanmar" in a research collaboration between Yangon Technological University, Myanmar, and The University of Tokyo, Japan, sponsored by the Japan Science and Technology Agency (JST) and the Japan International Cooperation Agency (JICA).

Keywords: BigGIS-RTX, CDR, OD trips, OD matrix, OD routes, OD links

1. Introduction

Big Data refers to data structures that are extremely large and complex in nature. Big Data is often characterized by 3Vs; Volume of data size, Velocity of real-time data streaming, and Variety of data formats such as text, image, video, and so on. These 3Vs were later extended to 5Vs by describing the Veracity of data content, especially Big Data coming from social media websites and other crowdsourcing sources, which have noise and abnormalities in the data, and the Value of the data, meaning the extent to which we can extract valuable information from Big Data. Sometimes, the Variability of Big Data mining and Visualization of Big Data analytical results are added to the Big Data category [1]. Big Data comes from sensor networks, weather stations, transaction machines, call details records (CDR), social media websites, and other crowdsourcing activities [2]. Moreover, the emergence of location-enabled mobile communication devices and other sensor network-based geospatial data acquisition systems has contributed a significant amount to Big Data, also known as Geospatial Big Data.

Big Data has been used for several decades in business management; for example, organizations discover their customer's behavior and product demands to understand their whole business process. Other examples include fraud detection and security analysis, disease dispersion and outbreak pattern analysis, and other scientific research applications. Owing to the large size of the data, complexity of the data structure, and links between data, the handling of Big Data will never be easy, and computational resources and analytical tools or software are required to turn this Big Data into valuable information. Nowadays, Big Data analysis is the heart of a business and can normally be done in either a cloud-based or a standalone PC computing environment or through interactive analysis of real-time data [3]. Apache Hadoop is an open-source cloud-based data-processing platform that is widely used by many organizations and institutions for Big Data processing. However, the use of cloud-based Big Data processing platforms is not suitable for developing countries, as they require an advanced network infrastructure, high-speed Internet connections, and knowledge



OD route generation module

Fig. 1. BigGIS-RTX functional menu and graphical user interface GUI (preliminary version).

of cloud-based computing.

Processing and analysis of Big Data is important for turning Big Data into useful information that can improve decision making in many planning processes and management issues. Moreover, its use by researchers in many academic fields including transportation and urban planning, human mobility and social interaction analysis, health, and other geospatial researches is increasing, due to the rich attribute information with geographical locations (longitude and latitude). Therefore, a geographical information system (GIS) embedded text-based geospatial Big Data research toolbox (BigGIS-RTX) is designed to act as a bridge between geospatial Big Data and end users (i.e. researchers and students) and can be run on a standalone computer without an Internet connection or knowledge of cloud computing. This research toolbox makes it possible to convert geospatial Big Data into meaningful information especially designed for CDR data to extract information on the dynamic population distribution and generate origin–destination (OD) trips, OD matrices, OD routes, and OD links in a user-defined time frame (e.g. trips between 06:00 and 09:00) and time window (e.g. continuous calls made within one or two hours). This article is composed of two sections. The first briefly introduces the BigGIS-RTX functional menu and graphical user interface (GUI); the second presents the case study results of dynamic population estimation and generation of OD trips, OD matrix, OD routes, and OD links from CDR data using BigGIS-RTX.

2. BigGIS-RTX GUI

The functional menu, shown in **Fig. 1** can be divided into two sub-menus: the General Module and CDR Module. The General Module provides the fundamental requirements of geospatial Big Data preprocessing such as data cleaning, formatting, finding and replacing strings, subsetting the data and merging files, and extraction by desired search conditions, rows, and columns from the data by using a structured query language (SQL). The CDR sub-module is especially designed for CDR data processing in transportation studies. This BigGIS-RTX is part of the main Geospatial Big Data Toolbox program, which is still under development and has been used in many research and development projects such as data-driven Web-GIS for a geospatially enabled person-trip data browser and space-time visualizer [4], correlation analysis between rainfall and geotagged Twitter data [5], a space-time multiple regression model for grid based population estimation [6], land cover weighted multitemporal grid square population estimation from CDR data [7], mapping the spatial distribution patterns of personal time spent based on trip purpose [8], human mobility patterns [9], and estimation of OD trips [10, 11]. **Fig. 2** shows the functional work flow of BigGIS-RTX in human mobility and transportation research using CDR data.

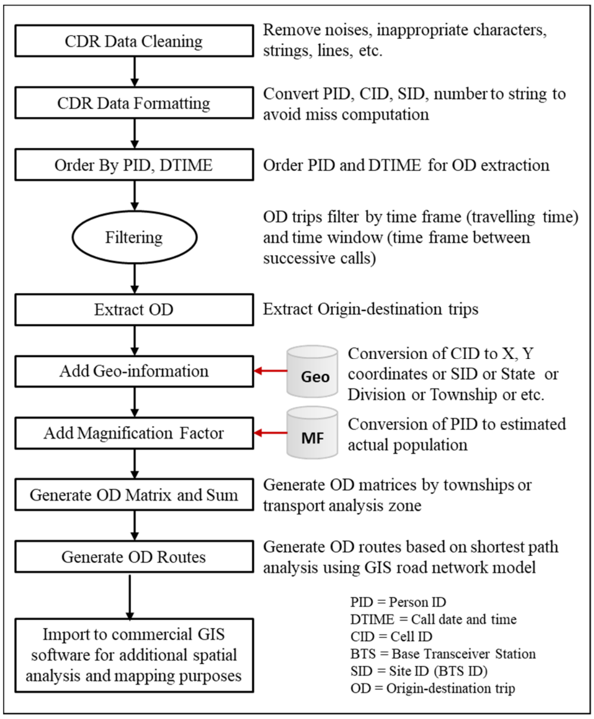


Fig. 2. BigGIS-RTX data processing work flow for Origin-destination trips extraction and route distance computing.

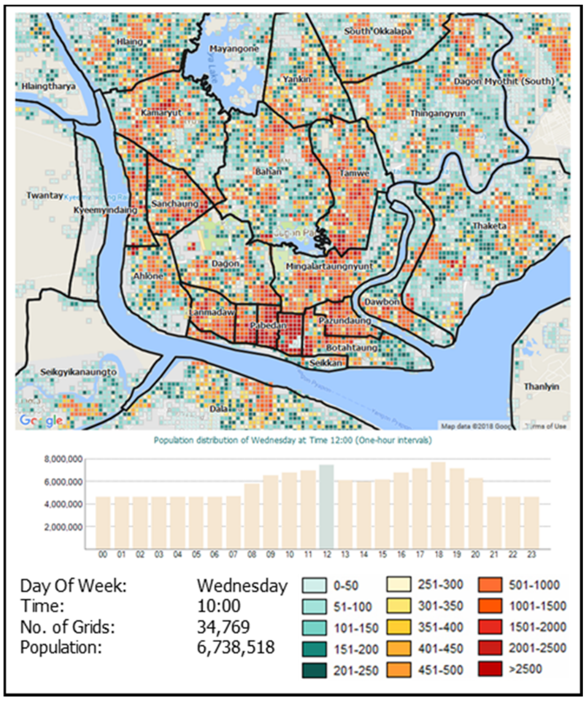


Fig. 3. Multitemporal grid square population estimated from CDR data and national census.

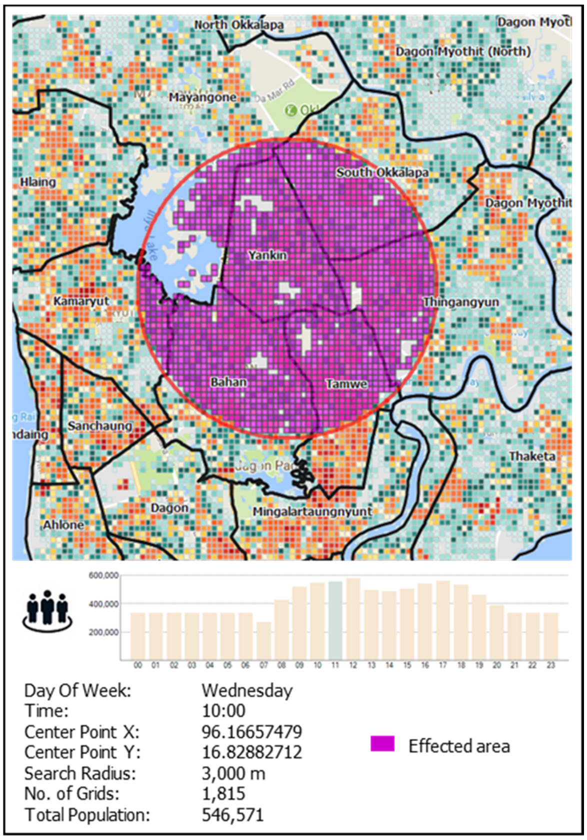


Fig. 4. Estimation of affected population in 3-km radius at user defined point such as epicenter when earthquake occurs.

3. Case Study Results Using BigGIS-RTX

3.1. Multitemporal Grid Square Population Estimation from Mobile CDR Data

This research aims to estimate the multitemporal grid square population in one-hour intervals from CDR data and to incorporate it with national census data. **Fig. 3** shows the multitemporal grid square population estimated from CDR data and census data. This data can be used for the estimation of the affected population in earthquake epicenters with a user-defined radius in kilometer (**Fig. 4**), affected population in flooded regions by specific distance buffered zone along the road or river (**Fig. 5**), user defined specific business zone or disaster area (**Fig. 6**) and so on. Moreover, a dynamic grid population dataset is essential for disaster management and emergency preparedness processes by integration with other GIS data layers.

3.2. Generation of OD Trips, Matrices, Routes, and Links

We used BigGIS-RTX to extract OD trips, OD matrices, OD routes, and OD links from CDR data. All this information is required for transportation planning and travel behavior studies in sustainable urban development. This tool provides GIS functions for the allocation of the cell-ID to latitude and longitude information, traffic analysis zones (TAZs), or administration units. Moreover, this tool can find the shortest routes and distance between OD pairs to analyze the average number of people travelling by time, distance, or trip; this information is useful for travel behavior and demand studies in transportation planning. **Fig. 7** shows a conceptual diagram of OD extraction

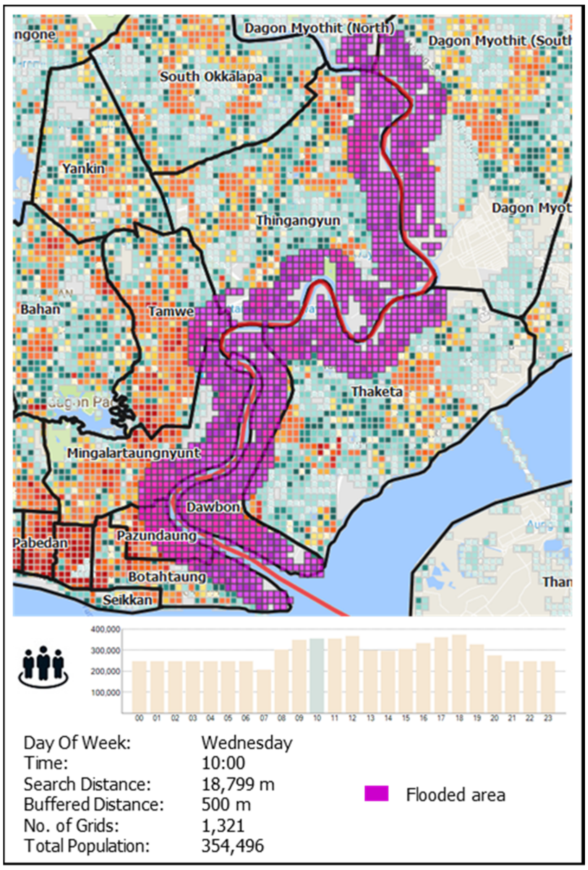


Fig. 5. Estimation of affected population in 500-m buffer zone along the river during flooding.

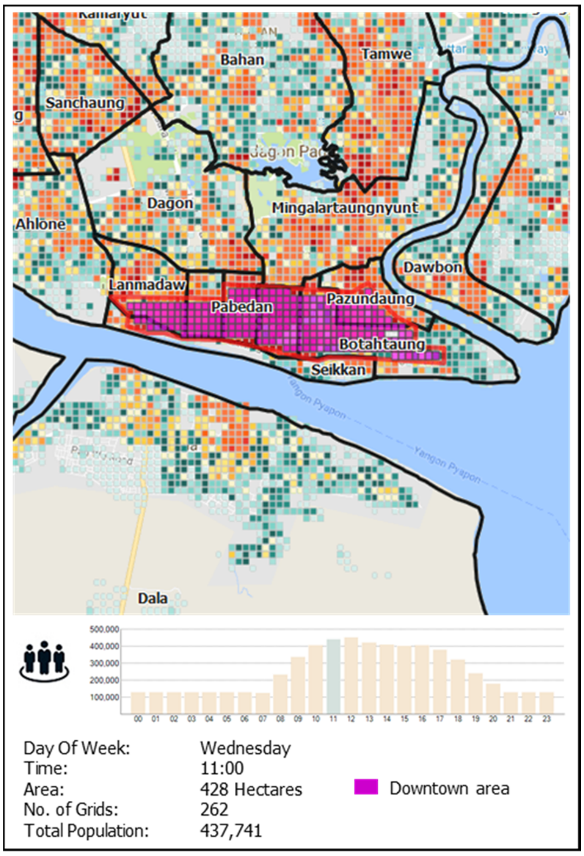


Fig. 6. Estimation of affected population by specific business zone or disaster area.

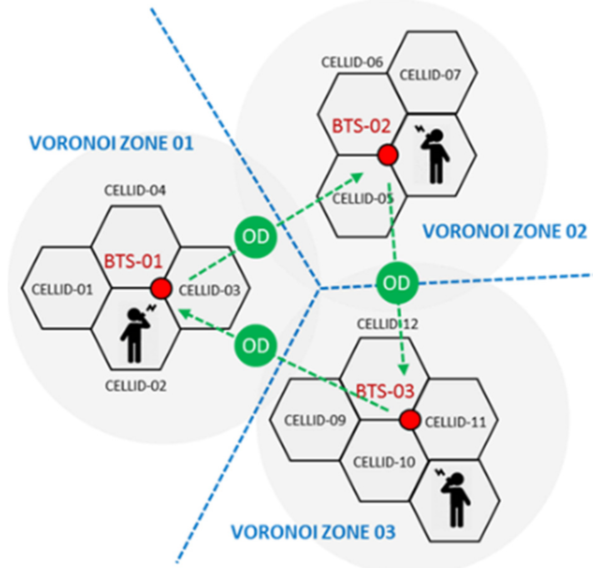


Fig. 7. Conceptual diagram of OD extraction from CDR data.

from CDR data.

This Toolbox can generate an OD matrix which measures the number of OD trips between the townships or TAZs (**Fig. 8**). OD matrices are essential for identifying flow magnitude and direction (**Figs. 9** and **10**).

Moreover, this Toolbox helps count the total number of unique PIDs passing the individual road segment (**Fig. 11**). This information is important for traffic volume and congestion analysis in transportation.

4. Future Development

Owing to the availability of geolocation information at the base transceiver station (BTS) level, trip generation is currently based on BTS locations. We are now developing a program to acquire cell-ID geolocations, which will enhance the accuracy of network distance measurement and identification of trip purposes by integration with the land use/cover ancillary GIS dataset.

Acknowledgements

We would like to thank MPT Myanma Post and Telecommunication for providing the one-week CDR data for this study. We also would like to acknowledge the Japan Science and Technology Agency (JST) and the Japan International Cooperation Agency (JICA) for funding and implementing the project entitled "Development of a Comprehensive Disaster Resilience System and Collaboration Platform in Myanmar" led by Professor Kimiro Meguro, International Center for Urban Safety Engineering at the Institute of Industrial Science, University of Tokyo, under the SATREPS Science and Technology Research Partnership for Sustainable Development program.

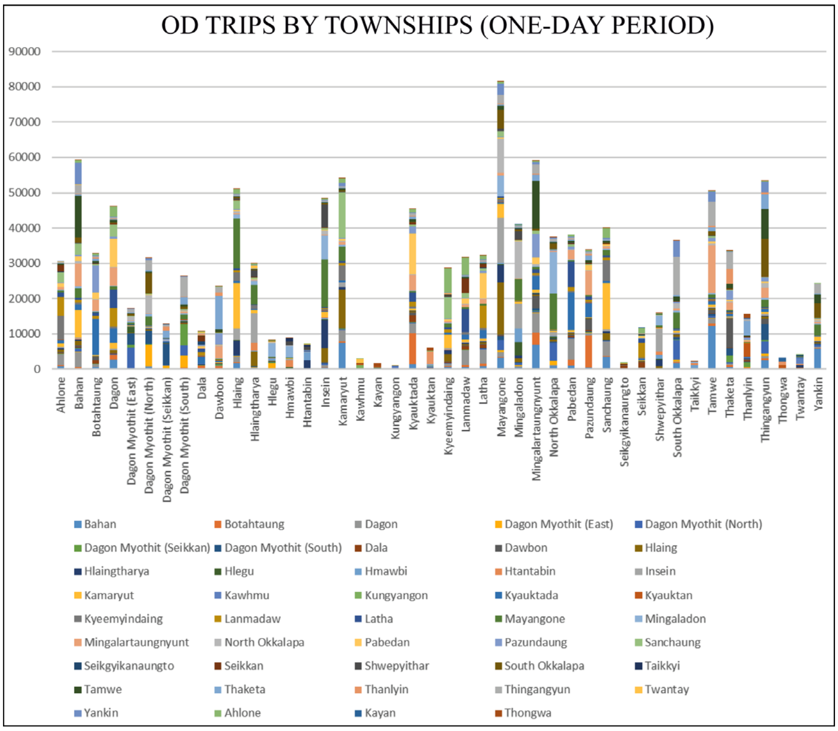


Fig. 8. Quantitative analysis of one-day OD trips between administrative zones.

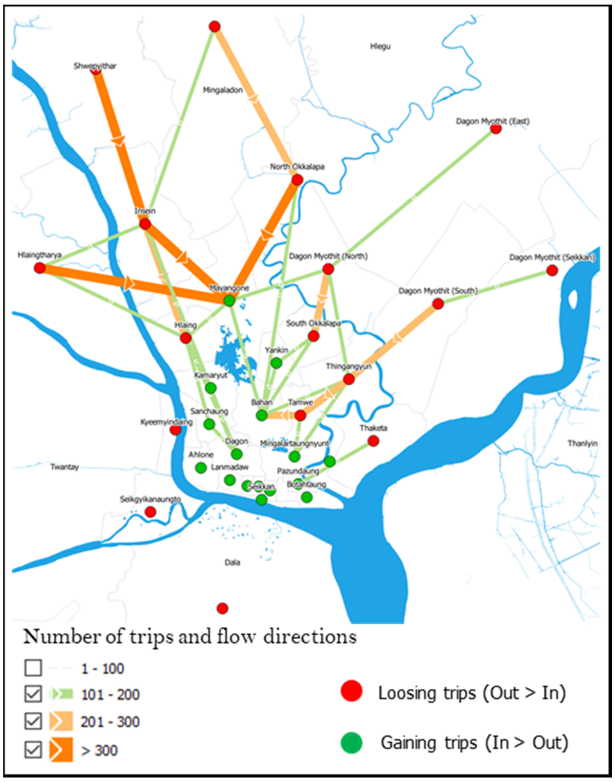


Fig. 9. OD magnitude and flow direction from CDR data at time between 05:00 – 09:00 (Morning peak hours).

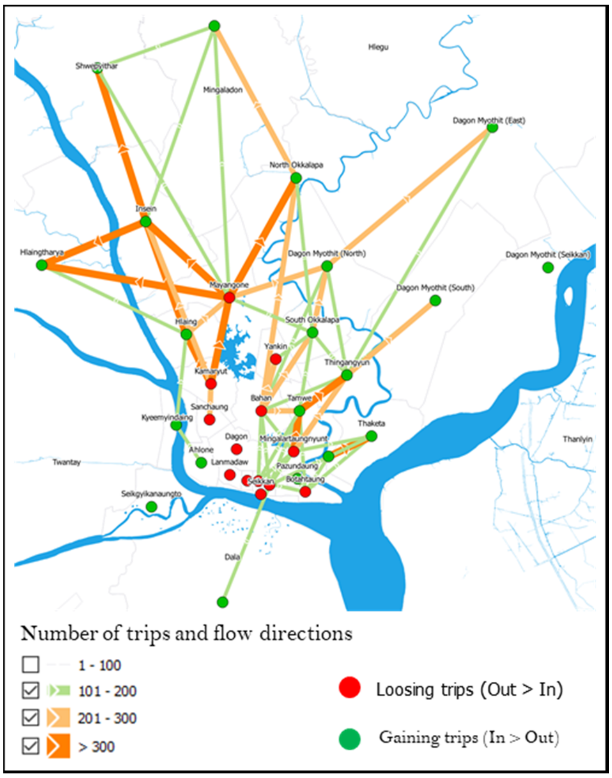


Fig. 10. OD magnitude (number of trips) and direction from CDR data at time between 17:00 – 21:00 (Evening peak hours).

References:

[1] M. Hilbert, "Big Data for development: A review of promises and challenges," Development Policy Review, Vol.34, No.1, pp. 135-174, 2015.

[2] C. Dobre and F. Xhafa, "Parallel programming paradigms and frameworks in Big Data era," Int. Journal of Parallel Programming, Vol.42, No.5, pp. 710-738, 2013.

[3] J. Lee and M. Kang, "Geospatial Big Data: Challenges and oppor-

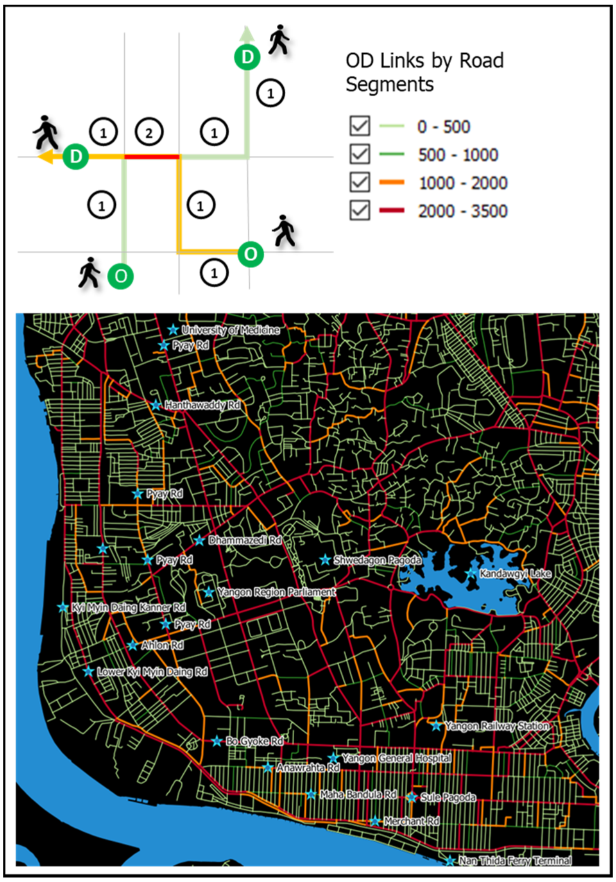


Fig. 11. OD links estimated from CDR data based on GIS road network data model (shortest path analysis).

tunities," Big Data Research, Vol.2, No.2, pp. 74-81, 2015.

[4] K. Lwin and Y. Murayama, "Person trip data browser and space-time visualizer," 2nd IEEE Int. Conf. on Spatial Data Mining and Geographical Knowledge Services, 2015.

[5] K. Lwin, K. Zettsu, and K. Sugiura, "Geovisualization and correlation analysis between geotagged Twitter and JMA rainfall data: Case of heavy rain disaster in Hiroshima," 2nd IEEE Int. Conf. on Spatial Data Mining and Geographical Knowledge Services, 2015.

[6] K. Lwin, K. Sugiura, and K. Zettsu, "Space–time multiple regression model for grid-based population estimation in urban areas," Int. Journal of Geographical Information Science, Vol.30, No.8, pp. 1579-1593, 2016.

[7] K. Lwin, Y. Sekimoto, and W. Takeuchi, "Land cover weighted mobile CDR data disaggregation for multitemporal grid based population estimation: A case of Yangon central business area, Myanmar," Monitoring of Global Environment and Disaster Risk Assessment from Space: the IIS Forum proceedings, Vol.25, pp. 47-50, 2017.

[8] K. Lwin and Y. Sekimoto, "Mapping the spatial distribution patterns of personal time spent based on trip purpose," Int. Journal of Applied Geospatial Research (IJAGR), Vol.9, No.2, 2018.

[9] D. Kyaing, K. Lwin, and Y. Sekimoto, "Human mobility patterns for different regions in Myanmar based on CDRs data," 3rd Int. Conf. on Civil Engineering Research (ICCER) and Regional Conf. in Civil Engineering (RCCE), 2017.

[10] D. Kyaing, K. Lwin, and Y. Sekimoto, "Estimation of trip generation in Yangon City by using CDRs data," Proc. of 12th Int. Conf. of Eastern Asia Society For Transportation Studies, Vol.11, 2017.

[11] T. Zin, D. Kyaing, K. Lwin, and Y. Sekimoto, "Estimation of Origin-Destination Trips by Using Big Data Source in Yangon," Journal of Disaster Research, Vol.13, No.1, 2018.

Name:
Ko Ko Lwin

Affiliation:
Assistant Professor, Dr., Institute of Industrial Science, The University of Tokyo

Address:
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

Brief Career:
2004 Senior Engineer, Vision Tech, Tsukuba
2010 Researcher, Division of Spatial Information Science, University of Tsukuba
2014 Researcher, National Institute of Information and Communications Technology (NICT)
2016 Assistant Professor, Institute of Industrial Science, The University of Tokyo

Selected Publications:
- K. Lwin, M. Hashimoto, and Y. Murayama, "Real-Time Geospatial Data Collection and Visualization with Smartphone," Journal of Geographic Information Systems, Vol.6, pp. 99-108, 2014.
- K. Lwin and Y. Murayama, "Modelling of Urban Green Space Walkability: Eco-friendly Walk Score Calculator," Computers, Environment and Urban Systems, Vol.35, No.5, pp. 408-420, 2011.

Academic Societies & Scientific Organizations:
- IEEE Advancing Technology for Humanity
- Urban and Regional Information System Association (URISA), US
- Japan Society of Photogrammetry and Remote Sensing (JSPRS)

Name:
Yoshihide Sekimoto

Affiliation:
Associate Professor, Institute of Industrial Science, the University of Tokyo

Address:
4-6-1, Komaba, Meguro-ku, Tokyo, Japan

Brief Career:
1999 Master of Engineering in Graduate School of Engineering, the University of Tokyo
2002 Ph.D. in Graduate School of Engineering, the University of Tokyo
2002-2007 Working as Researcher at National Institute of Land, Infrastructure and Management, Ministry of Land, Infrastructure, Transport and Tourism
2007-2013 Working as Project Associate Professor at Center for Spatial Information Science, the University of Tokyo
2013- Working as Associate Professor at Institute of Industrial Science, the University of Tokyo

Selected Publications:
- "PFLOW: Reconstruction of people flow by recycling large-scale fragmentary social survey data," IEEE Pervasive Computing, 2011

Academic Societies & Scientific Organizations:
- Institute of Electrical and Electronics Engineers, Inc. (IEEE)
- GIS Association of Japan (GISA)
- Japan Society of Civil Engineers (JSCE)
- Japan Society of Traffic Engineers (JSTE)

Name:
Wataru Takeuchi

Affiliation:
Associate Professor, Institute of Industrial Science, The University of Tokyo

Address:
Bw-602, 4-6-1, Komaba, Meguro, Tokyo 153-8505, Japan

Brief Career:
2004 Joined Institute of Industrial Science, The University of Tokyo
2007-2009 Visiting Scholar, Asian Institute of Technology, Thailand
2010-2012 Director, Bangkok office, Japan Society for Promotion of Science (JSPS)
2017- Senior Policy Analyst, Council for Science, Technology and Innovation (CSTI), Cabinet Office of Japan

Selected Publications:
- W. Takeuchi, T. Hirano, and O. Roswintiarti, “Estimation of ground water table at forested peatland in Central Kalimantan, Indonesia,” Tropical Peatland Ecosystems, Springer Verlag, 2016.
- W. Takeuchi, T. Nakano, and Y. Yasuoka, “Estimation of methane emission from west Siberian lowland with sub-pixel land cover characterization between MODIS and ASTER,” Land remote sensing and global environmental change: NASA’s EOS and the science of ASTER and MODIS, Springer Verlag, 2011.

Academic Societies & Scientific Organizations:
- Japan Society for Photogrammetry and Remote Sensing (JSPRS)
- Remote Sensing Society of Japan (RSSJ)
- American Geophysical Union (AGU)