IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

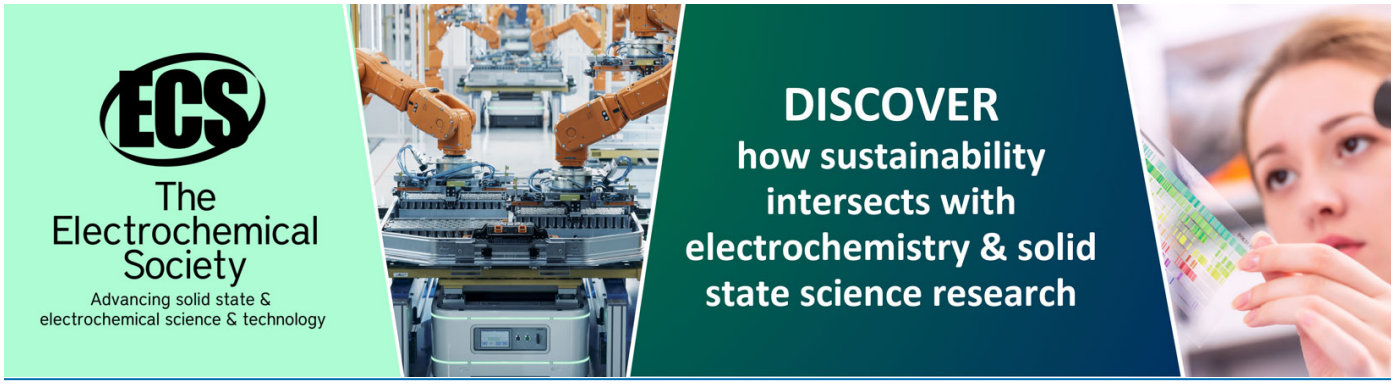
Application of geodata for operational study of population placement and movement

To cite this article: A N Vorobyev 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **629** 012003

View the article online for updates and enhancements.

You may also like

- THINGS: THE H i NEARBY GALAXY SURVEY
Fabian Walter, Elias Brinks, W. J. G. de Blok et al.
- Variational quantum compiling with double *Q*-learning
Zhimin He, Lvzhou Li, Shenggen Zheng et al.
- GIS approaches to creating maps based on vegetation indices for forest management
Y Y Savchenko, O G Goleva, I A Korchagina et al.



This content was downloaded from IP address 3.133.109.30 on 03/05/2024 at 13:14

IOP Conf. Series: Earth and Environmental Science **629** (2021) 012003 doi:10.1088/1755-1315/629/1/012003

Application of geodata for operational study of population placement and movement

A N Vorobyev

Sochava Institute of Geography SB RAS, Irkutsk, Russia

E-mail: tore12@yandex.ru

Abstract. The aim of this article is to develop a methodology for compiling population density and mobility maps based on the data (geographic data) obtained from mobile operators in Irkutsk. Before its development, Russian and world experience in compiling maps based on mobile phones data was analysed. To achieve the aim, the author uses the data received from mobile operators (CDR), with which it was possible to identify information about the location of the device at any time. Based on the geographical location of base stations, one can analyse the density of users at a particular time and place. Based on the results of spatial data analysis, we obtained the signal density at a particular time. The results obtained now clearly demonstrate that geographic data from user devices and geographic information systems represent a very effective symbiosis of technologies, which serves for solving geographical problems. Such studies assess the general mobility of population for a particular region. Data on the location of a mobile phone can reveal the spatiotemporal distribution of urban residents at macro- and meso-levels in administrative districts of the city.

1. Introduction

In the Russian-speaking sphere, "Big Data" is a literal translation of the English expression "big data", which was drawn from information technology (IT) and settled in almost all spheres of human life. For the first time, in the presentation made in 1998, John Mashey, who was at that time the chief scientist of 'Silicon Graphics' company, used the term Big Data. However, then the term was not widely used, since Mashey predicted future data growth addressing a narrow circle of colleagues [1]. The phrase Big Data gained its current popularity after being published in the journal Nature in 2008. In September 2008, a special issue of Nature "How can technologies that open up opportunities for working with large amounts of data affect the future of science?" was published. The issue contained materials on the phenomenon of explosive growth in the volume and variety of processed data and technological prospects in the paradigm of a probable jump "from quantity to quality"; the term was proposed by analogy with the metaphors "big oil" and "big ore" that are common in the business English-speaking environment. "Big data" means a set of technologies for searching, collecting, analyzing, storing, processing, etc. significant amounts of information. The main differences between big data and other data can be described as "three V" – volume, velocity and variety [2]. In many cases, big data includes a direct or an indirect reference to a location on the Earth and may then be referred to as "big geospatial data".

The aim of the study is the use of geodata in the study of location and spatial movement of the population as well as analysis of demographic processes.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

2. Materials and methods

Geodata is spatiotemporal data reflecting the properties of objects, processes and phenomena taking place on the Earth. They contain information about objects, forms of territory and infrastructures on the earth's surface, and, as an essential element, they must necessarily contain spatial relationships [3].

In the modern world, many devices (phones, navigators, smartwatches, etc.) collect huge amounts of data (geodata). In the recent past, the collection of geospatial data was complicated by technical capabilities. To collect information, technically complex, cumbersome and expensive devices were required, where the measurement process itself was laborious work and required highly qualified specialists. However, the rapid growth of information technology in household devices, such as smartphones and others, used by many people, has made it possible to move to a qualitatively new level of geodata collection. Today's devices are capable of acquiring geospatial information at an unprecedented level in terms of significantly improved accuracy, temporal resolution and thematic detail. Mobile devices are small, easy to use and able to receive data even without the subscriber being aware. Smartphones and mobile devices are ubiquitous, which enables continuous collection of user information through sensors and GPS receivers. The capabilities of the tracking sensors also extend to vehicles that record the location of the vehicle in real time. A clear illustration is the display of the situation on the road in various car navigators (Google, Yandex, etc.) on the example of Irkutsk.

Billions of transactions are carried out throughout the world by bankcards and contactless payments with smartphones, each of which also leaves its own digital trail.

The main feature of geodata is the accurate reflection of the subscriber's location at a particular moment and the ability to obtain exact coordinates as well as the probability of restoring the chronology of the subscriber's movements.

Another example of all modern technologies' combination for working with "big data" and the use of human resources is the free web-based mapping service OpenStreetMap. Any registered user, who becomes a supplier of "big data" in the form of GPS tracks, aerial photographs, video recordings, etc., can join the creation of the map. Such users freely participate in the collection of data, being, in fact, volunteers. In this regard, a new term arose in cartography and geoinformatics: “Volunteered Geographic Information (VGI)" [4].

One example of modern technologies keeping pace with time is the concept of the self-isolation index introduced by Yandex and associated with the COVID-19 virus pandemic. The index is an integral indicator based on all services of the company. When calculating the self-isolation index in Yandex, the number of people on a weekday at rush hour is taken as zero. If the activity is similar to the one that happens at night, then everyone stays at home and the self-isolation index is 5.

At the same time, the limits from 0 to 2.4 are marked in red (there are many people in the streets); yellow – 2.5-3.9 (there are relatively many people in the streets) and green – 4-5 points (there are almost no people in the streets). With the help of these data, it is possible to track how the population complies with the self-isolation regime introduced by the Government of the Russian Federation.

In this article, we will look at receiving data using a mobile phone. Interest in the use of the data obtained from mobile phones is growing rather rapidly due to the fact that phones are the most common gadgets among all segments of the population (children, adults and elderly users), and they also appear with more new functions.

3. Results and discussion

To obtain data on the location and movement of mobile phone users, you can use the data available from telecommunications companies. Telecommunications companies are organizations that provide customers with mobile or fixed telephone services, high-speed Internet access and cable television.

Telecommunications companies now have sufficient information about their users to compete on an equal footing with social networks [5]. With appropriate processing, this information can provide a huge body of knowledge that cannot be obtained in any other way. After analyzing geospatial data, one can get accurate and, importantly, operational information and understand, what processes are taking place in the city during the day. All information is accumulated in the Call Detail Record

(CDR) [6] that includes information about the location of the device at each moment, a call log, including information about another subscriber, and data on Internet sessions. Concerning SMS, usually without prior authorization for wiretapping, the operator has the right (and is obliged) to save metadata only: the time of sending, the size of the message and the addressee. The content of the messages themselves (let alone voice calls) is not saved. In July 2016, Russia passed the Federal Law of 6 July 2016 № 374-FZ «On Amendments to the Federal Law «On Combating Terrorism» and certain legislative acts of the Russian Federation to establish additional measures to counter-terrorism and ensure public safety» («Yarovaya law»), which obliges cellular operators to store metadata for three years. Additionally, from October 1, 2018, operators are required to store text, voice, video and other messages of users for at least 30 days (but less than six months).

In our opinion, using mobile phone data as an alternative source of population distribution will significantly increase the accuracy of mapping. With the rapid development of information and communication technologies (ICT), mobile phone data is becoming an important source for studies of population distribution and movement (commuting) of urban residents [7]. Despite obvious advantages, in our country, studies of population distribution and movement using data from mobile phones as a source of information are not as popular as among foreign colleagues.

Meanwhile, a huge number of mobile phone base stations with corresponding user data provide information on the spatial heterogeneity of users. However, the data obtained from traditional sources (population censuses and statistical compilations) have clear boundaries, usually administrative and territorial. Population data associated with mobile phone base stations does not have a hard-to-reach service area.

In such studies, the main difficulty lies in obtaining data from mobile operators for a certain period in order to track dynamic effects. In Russia, the largest mobile operators are MTS, Beeline, Megafon, and Tele2.

The data is pre-processed to exclude the information related to the privacy of subscribers. Mobile operators declare that all data is impersonal and only includes the age and geolocation of the subscriber [8].

The basic format is a multi-field table marked with a user ID. Subsequently, the data from base stations within one month (January-February) were divided into three periods: working hours (Monday to Friday) from 7-00 to 19-00, non-working hours - from 19-00 to 7-00, weekends and holidays and for the entire period. [9] We believe that the data for the late January - early February are the most consistent with the objectives of the study, since the winter months and the absence of long weekends contribute to a more homogeneous population movement.

Temporal and spatial locations are associated with base stations through a user identifier (ID). Based on the geographic location of the base stations, it is possible to analyze the density of users at a specific time at a certain point. With a dense distribution of base stations in central urban areas with heavy population, especially in metropolitan areas, the subscriber location error can be within a few hundred meters. Subsequently, the spatial distribution of base stations is plotted in QGIS (free open source geographic information system), and a relational table of different base stations with specific IDs is created. Moreover, around each base tower, we get the number of users on working and non-working days.

Compared to (traditional, generally accepted) data sources, such as official census statistics, mobile phone location data has clear advantages in terms of statistical accuracy and timeliness. Thus, data on the location of a mobile phone can really reveal the spatial and temporal distribution of urban residents at the macro- and meso-levels in the administrative districts of the city [10].

Further, the positions of the base station are processed by the spatial analysis tools "Voronoi polygons" in QGIS [11]. Voronoi's polygon is the influence of one point on some space. Voronoi polygons are created based on mobile base stations network; the number of phone outputs to the network for a certain period (total, weekends and working days) is divided by the area within a certain polygon. Ultimately, we get the density of signals in a certain period.

4. Conclusions

Thus, we indicate that GIS and data (geodata) from user devices are effective tools for solving geographic problems. The scale of the study of data (geodata) will inevitably increase in the coming years. Human (user) mobility research is a powerful tool for the analysis of individual patterns of human movement, facilitating the operational study of population placement and movement. Such studies assess the general mobility of the population for a particular region and movement on a street scale, etc., which is impossible to do, relying only on official sources of information (statistical compilations and population censuses). However, they are unlikely to replace traditional research on human mobility, which provides a deeper understanding of socioeconomic characteristics. A rational combination of traditional [12] and new methods can improve the quality, accuracy and efficiency of knowledge about the localization and mobility of the population.

Acknowledgments

This work was carried out with the financial support from the Russian Federal Property Fund within the framework of the scientific projects No. 19-55-44020 Mong_t and No. 17-29-05043 ofi_m

References

[1] Chernyak L A Fresh look at Big Data 2013 *Open Systems* **7** 48-51 Income accessed online on 16th September 2020 via https://www.osp.ru/os/2013/07/13037355/

[2] Agrawal D *et al*. 2011 Challenges and opportunities with Big Data *Cyber Center Technical Reports* **1** 16 Income accessed online on 16th September 2020 via http://docs.lib.purdue.edu/cctech/1

[3] Tsvetkov V Ya and Domnitskaya E V 2008 Geodata as the basis of digital modeling *Modern Science-Intensive Technologies* **4** 100-1

[4] Nyrtsov M V and Nyrtsova T P 2016 Big Data in cartography. smart mapping: the future or technological change *News of Higher Educational Institutions. Geodesy and Aerial Photography* **5** 42-5 (in Russian)

[5] Deville P, Linard C, Martin S, Gilbert M, Stevens F R, Gaughan A E, Blondel V D and Tatem A J 2014 Dynamic population mapping using mobile phone data *Proc. of the National Academy of Sciences* **111**(45) 15888-93 doi:10.1073/pnas.1408439111

[6] Cumbley R and Church P 2013 "Is “Big Data” creepy?" *Linklaters LLP, Computer Law & Security Review* **29**(5) 601-9

[7] Babkin R A 2019 Spatial dynamics of the Moscow agglomeration *Geographical Research of Siberia and Adjacent Territories* (Irkutsk: Publishing House Sochava Institute of Geography SB RAS) pp 22-5 (in Russian)

[8] Portela J N and Alencar M S 2008 Cellular coverage map as a Voronoi diagram *Journal of Communication and Information Systems* **23(**1) 22-31

[9] Wang W, Pei T, Chen J, Song C, Wang X, Shu H, Ma T and Du Yu 2019 Population Distributions of Age Groups and Their Influencing Factors Based on Mobile Phone Location Data: A Case Study of Beijing, China *Sustainability* **11** 7033 doi:10.3390/su11247033

[10] Gariazzo C and Pelliccioni A 2019 A multi-city urban population mobility study using mobile phone traffic data *Applied Spatial Analysis and Policy* **12** 753-71

[11] Makhrova A G, Babkin R A and Kazakov E E 2020 the dynamics of the day and night population as an indicator of structural and functional changes in the territory of the city in the zone of influence of the moscow central ring using data from mobile operators *Outlines of Global Transformations: Politics, Economics, Law* **13**(1) 159-79 (in Russian) doi:10.23932/2542-0240-2020-13-1-9

[12] Vorobyev A N 2019 The mapping population in a sparsely populated region: a case study the Irkutsk region *IOP Conf. Series: Earth and Environmental Science* **381** 012095 doi:10.1088/1755-1315/381/1/012095