

Использование геоданных для оперативного изучения населения

А. Н. Воробьев ✉

Институт географии им. В. Б. Сочавы Сибирского отделения РАН, Российская Федерация
(664033, г. Иркутск, ул. Улан-Баторская, 1)

Аннотация: Цель данной статьи – разработка методики составления карт плотности и мобильности населения, опираясь на данные (геоданные) сотовых операторов г. Иркутска. Перед ее разработкой проанализирован российский и мировой опыт составления карт по данным мобильных телефонов.

Материалы и методы. Для достижения цели автором используется информация, полученная от мобильных операторов (CDR), с помощью которой удалось выявить сведения о местоположении устройства в каждый момент времени. Зная географическое местоположение базовых станций, можно проанализировать плотность пользователей в конкретное время в определенной точке. На основе результатов пространственного анализа данных нами получена плотность сигналов в определенный промежуток времени.

Результаты и обсуждение. Полученные результаты в настоящее время наглядно демонстрируют, что геоданные с пользовательских устройств и геоинформационные системы представляют очень эффективный симбиоз технологий для решения географических задач. Такие исследования позволяют оценить общую мобильность населения для конкретного региона. Данные о местонахождении мобильного телефона способны действительно выявить пространственно-временное распределение городских жителей на макро- и мезоуровнях в административных районах города.

Заключение. Используя официальные источники информации (статистические сборники и переписи населения), невозможно оперативно рассмотреть передвижение населения на локальном уровне. Однако, при всех своих неоспоримых плюсах в ближайшее время геоданные вряд ли заменят традиционные методы изучения мобильности людей. Только рациональное сочетание традиционных и новых методов способно дать качественное изменение в точности и оперативности размещения населения.

Ключевые слова: большие данные, геоданные, мобильный телефон, плотность населения, мобильность населения.

Источник финансирования: Работа выполнена в рамках программы НИР Института географии им. В. Б. Сочавы СО РАН (проект № 0347-2016-0004).

Для цитирования: Воробьев А. Н. Использование геоданных для оперативного изучения населения // Вестник Воронежского государственного университета. Серия: География. Геоэкология, 2021, № 2, с. 49-54. DOI: <https://doi.org/10.17308/geo.2021.2/3447>

ВВЕДЕНИЕ

В русскоязычной среде «Большие данные» является буквальным переводом английского выражения «big data» перекочевавшего из информационных технологий (ИТ) и обосновавшегося практически во всех сферах жизни людей. Известно, что впервые в своей презентации, сделанной в 1998 году, термин Big Data использовал Джон Мэши – главный ученый компании Silicon Graphics. Однако тогда термин не получил широкого распространения, поскольку Мэши предсказывал будущий рост данных, адресуясь к узкому кругу

коллег. Свою нынешнюю популярность словосочетание Big Data обрело после публикации в журнале Nature в 2008 году. В сентябре 2008 года был издан специальный выпуск Nature «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных?». В номере были собраны материалы о феномене взрывного роста объемов и многообразия обрабатываемых данных и технологических перспективах в парадигме вероятного скачка «от количества к качеству». Термин был предложен по аналогии с расхожими в деловой англоязычной

© Воробьев А. Н., 2021

✉ Воробьев Александр Николаевич, e-mail: Tore12@yandex.ru



Контент доступен под лицензией Creative Commons Attribution 4.0 License.

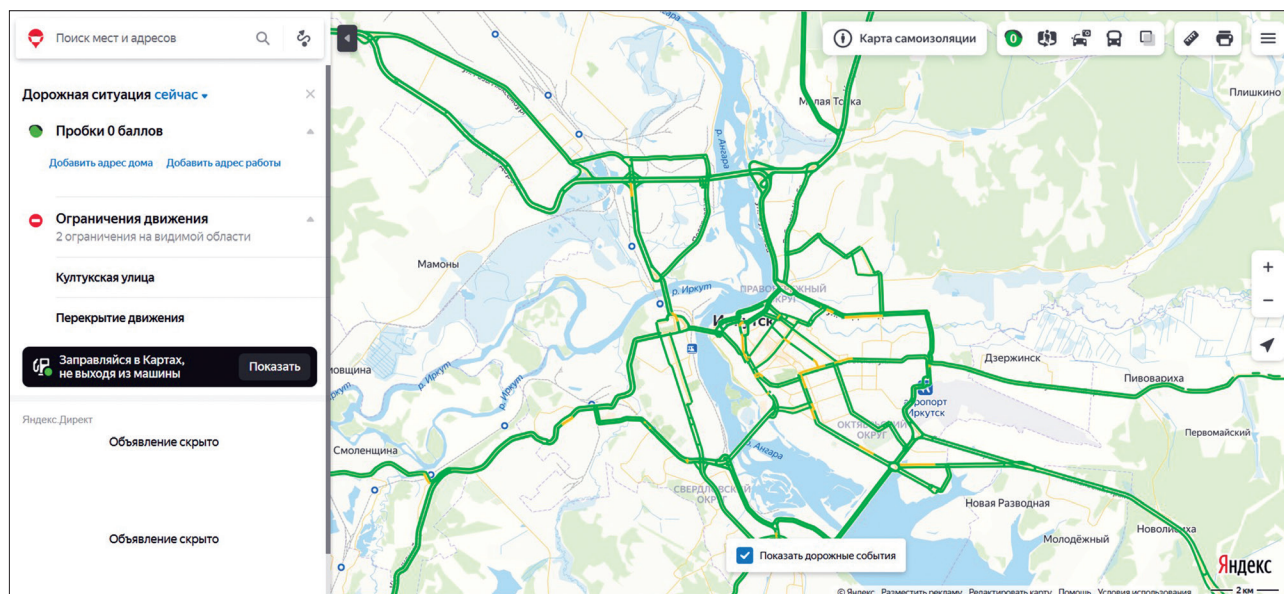


Рис. 1. Яндекс пробки г. Иркутска, 14.04.2020 [4]
 [Fig. 1. Yandex traffic jams in Irkutsk, 14.04.2020 [4]]

среде метафорами «*большая нефть*», «*большая руда*» [2]. «*Большие данные*» подразумевают совокупность технологий по поиску, сбору, анализу, хранению, обработке и т. п. значительных объёмов информации. Основные отличия больших данных от других можно охарактеризовать как «три V» (volume, velocity, variety) – объём, скорость, многообразие. Во многих случаях большие данные включают в себя прямую или косвенную ссылку на местоположение на Земле и могут затем называться «*большими геопространственными данными*».

Цель исследования – применение геоанных в изучении размещения, пространственного перемещения населения и анализе демографических процессов.

МАТЕРИАЛЫ И МЕТОДЫ

Геоанные – пространственно-временная информация, отражающая свойства объектов, процессов и явлений, происходящих на Земле. Они содержат информацию о предметах, формах территории и инфраструктурах на поверхности Земли, причем как существенный элемент в них должны обязательно присутствовать пространственные отношения.

В современном мире множество устройств (телефоны, навигаторы, смарт-часы и др.) занимаются сбором огромных массивов данных (геоанных). В недалеком прошлом сбор геопространственных данных был осложнен техническими возможностями. Для сбора информации требовались технически сложные, громоздкие и дорогие

устройства, где сам по себе процесс измерения был трудоемкой работой и требовал высокой квалификации специалистов. Однако, стремительный рост информационных технологий в бытовых устройствах, таких как смартфоны, используемые многими людьми, позволил перейти на качественно новый уровень сбора геоанных. Современные устройства способны получать геопространственную информацию на беспрецедентном уровне в отношении значительно улучшенной точности, временного разрешения и тематической детализации. Мобильные устройства малы, просты в обращении и способны получать данные даже без ведома абонента. Смартфоны и мобильные устройства повсеместны, что позволяет непрерывно собирать информацию о пользователях с помощью датчиков и GPS приемников. Возможности датчиков слежения также распространяются на автомобили, фиксирующие расположение транспортного средства в режиме реального времени. Наглядной иллюстрацией может служить отображение ситуации на дороге в различных автомобильных навигаторах (Google, Yandex и др.) на примере города Иркутска (рис. 1).

Миллиарды транзакций осуществляется по всему миру при помощи банковских карт и бесконтактных платежей с использованием смартфонов, каждая из которых тоже оставляет свой цифровой след.

Основная особенность геоанных – точное отражение местоположения абонента в конкретный момент и возможность получить точные координаты, а также вероятность восстановления хронологии передвижения.

Другим примером сочетания в себе всех современных технологий работы с «большими данными» и использования человеческих ресурсов является свободный веб-картографический сервис OpenStreetMap. К созданию карты может подключиться любой зарегистрированный пользователь, который становится поставщиком «больших данных» в виде GPS-треков, аэроснимков, видеозаписей и др. Такие пользователи добровольно участвуют в сборе информации, являясь, по сути, волонтерами. В связи с этим в картографии и геоинформатике возник новый термин «Добровольческая (волонтерская) географическая информация Volunteered Geographic Information (VGI)».

Еще одним примером современных технологий, идущих в ногу со временем, можно считать понятие индекса самоизоляции, который ввела в обиход компания Яндекс, связанный с пандемией вируса covid-19. Индекс является интегральным показателем, опирающийся на все сервисы компании (рис. 2). При расчете индекса самоизоляции в Яндексе за нулевую отметку принимается количество людей в будний день в час пик. Если же активность похожа на ту, что бывает в ночные часы, то все сидят по домам и индекс самоизоляции равен 5.

При этом красным цветом отмечаются пределы от 0 до 2,4 (на улицах очень много людей); желтым – 2,5-3,9 (на улицах много людей) и зеленым – 4-5 баллов (на улицах нет почти никого). При помощи этих данных можно отследить, насколько население соблюдает режим самоизоляции, введенный Правительством РФ.

В статье мы рассмотрим получение данных с использованием мобильного телефона. Интерес к использованию данных, получаемых от мобильных телефонов, растет довольно быстро. Причина – телефоны – наиболее распространенный гаджет среди всех слоев населения (дети, взрослое население, пожилые пользователи) и все с большим количеством функций.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для получения данных о локализации и перемещениях пользователей мобильных телефонов можно воспользоваться данными, имеющимися у телекоммуникационных компаний. **Телекоммуникационные компании** – организации, предоставляющие клиентам услуги мобильной или стационарной телефонной связи, скоростного доступа в интернет и кабельного телевидения. В настоящее время у них имеется такой объем информации о пользователях, что они могут на равных конкурировать с социальными сетями. При соответствующей обработке эта информация может дать огромный массив знаний, который не получить иным образом. Проанализировав геопространственные данные, можно получить точную и, что немаловажно, оперативную информацию и понять какие процессы протекают в городе в течение дня. Вся информация аккумулируется в запись сведений о звонке - Call detail record (CDR) [5], куда включены сведения о местоположении устройства в каждый момент времени, журнал звонков, включая информацию о другом абоненте, и данные о сессиях выхода в интернет.

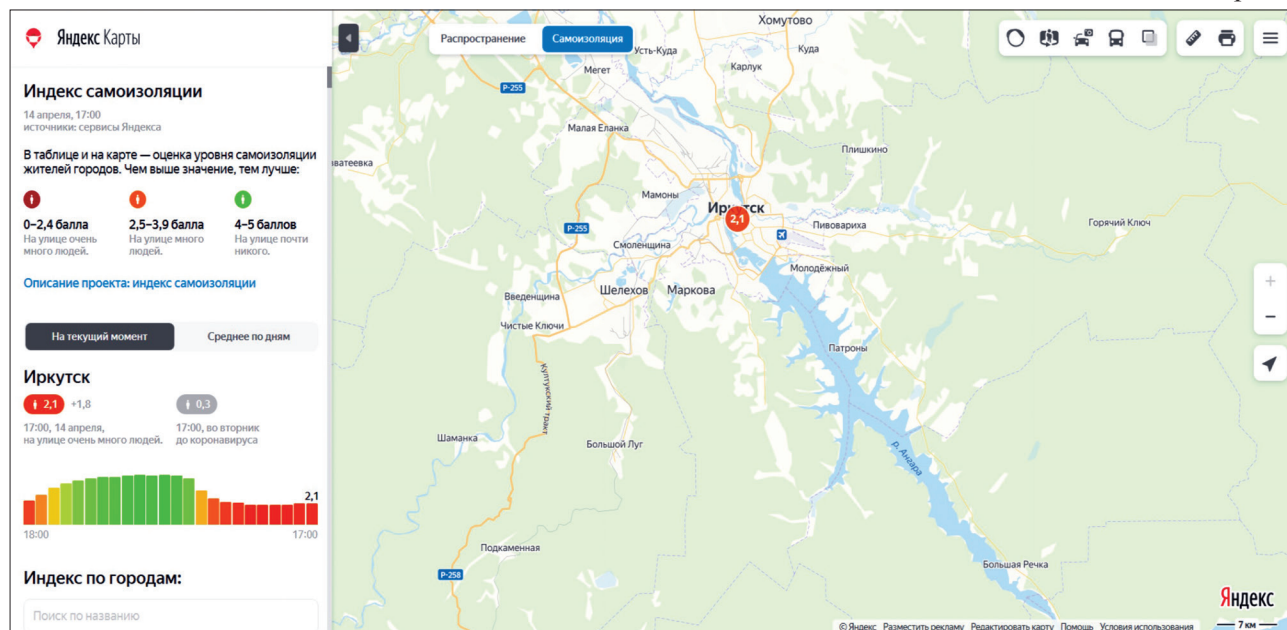


Рис. 2. Индекс самоизоляции г. Иркутска, 14.04.2020 [4]
[Fig. 2. Self-isolation Index of Irkutsk, 14.04.2020 [4]]

Что касается SMS, то обычно без предварительной санкции на прослушивание оператор имеет право (и обязан) сохранять лишь метаданные: время отправки, размер сообщения и адресат. Содержимое самих сообщений (а тем более голосовых звонков) не сохраняется. В июле 2016 года в России был принят закон Яровой, обязывающий операторов сотовой связи в течение трех лет хранить метаданные. С 1 октября 2018 года операторы обязаны хранить в течение как минимум 30 суток (но не более шести месяцев) текстовые, голосовые, видео- и другие сообщения пользователей [1].

Использование данных мобильного телефона в качестве альтернативного источника распределения населения значительно, по нашему мнению, повысит точность картографирования. С быстрым развитием информационно-коммуникационных технологий (ИКТ), данные мобильных телефонов становятся важным источником для исследований распределения населения и перемещения (маятниковая миграция) городских жителей. Несмотря на явные преимущества, в нашей стране исследования распределения и перемещения населения с применением данных с мобильных телефонов в качестве источника информации не столь популярно как у иностранных коллег.

Между тем, огромное количество базовых станций мобильных телефонов с соответствующими данными пользователей предоставляют сведения о пространственной неоднородности пользователей. Однако, данные, полученные из традиционных источников (переписи населения, статистические сборники), имеют четкие границы, как правило административно-территориальные. Данные о населении, связанные с базовыми станциями мобильных телефонов, не имеют жесткой границы зоны обслуживания.

При таких исследованиях основная сложность состоит в получении данных мобильных операторов за определенный период, чтобы отследить динамические эффекты. В России крупнейшие операторы мобильной связи – МТС, Билайн, Мегафон, Tele2.

Данные проходят предварительную обработку, чтобы исключить информацию, связанную с конфиденциальностью абонентов. Операторы сотовой связи заявляют, что все данные обезличены и включают в себя только возраст и геолокацию абонента.

Основной формат – это многопольная таблица, помеченная идентификатором пользователя. В дальнейшем, данные с базовых станций в течение

одного месяца (январь-февраль) были разделены на три периода времени. Рабочее время (с понедельника по пятницу) с 7-00 до 19-00, нерабочее – с 7-00 до 7-00 плюс выходные дни и праздничные и за весь период. Данные за конец января – начало февраля, мы считаем, больше всего соответствуют задачам исследования, так как зимние месяцы и отсутствие длительных выходных способствуют более однородному перемещению населения.

Посредством идентификатора пользователя (ID) временные и пространственные местоположения связаны с базовыми станциями. Полагаясь на географическое местоположение базовых станций, можно проанализировать плотность пользователей в конкретное время в определенной точке. При густом распределении базовых станций в центральных городских районах с плотным населением, особенно в мегаполисах, погрешность местоположения абонента может находиться в пределах нескольких сотен метров. Впоследствии в QGIS (свободная геоинформационная система с открытым кодом) наносится пространственное распределение базовых станций, и создается реляционная таблица различных базовых станций с определенными ID. Кроме того, вокруг каждой базовой вышки мы получаем количество пользователей в рабочие и нерабочие дни.

По сравнению с традиционными и общепринятыми источниками информации, такими как официальная статистика переписи населения, данные о местонахождении мобильных телефонов обладают очевидным преимуществом в статистической точности и своевременности. Таким образом, данные о местонахождении мобильного телефона способны действительно выявить пространственно-временное распределение городских жителей на макро- и мезоуровнях в административных районах города.

В дальнейшем позиции базовой станции обрабатываются инструментами пространственного анализа «полигоны Вороного» в QGIS. Полигон Вороного — это влияние одной точки на какое-то пространство. Полигоны Вороного создаются на основе сети базовых станций мобильной связи, количество выходов телефонов в сеть за определенный период (общее, выходные, рабочие дни) делится на площадь в пределах определенного полигона. Тем самым мы получаем плотность сигналов в определенный промежуток времени.

ЗАКЛЮЧЕНИЕ

Таким образом, наглядно демонстрируется, что ГИС- и данные (геоданные) с пользовательских устройств представляют из себя эффективный

инструмент для решения географических задач. Масштабы изучения данных (геоданных) неминуемо в ближайшие годы увеличатся. Исследования мобильности людей (пользователей) представляет мощный инструмент для анализа индивидуальных моделей передвижения человека, способствующий оперативному изучению размещения и перемещению населения.

Такие исследования позволяют оценить общую мобильность населения для конкретного региона. Изучить передвижения человека в масштабе улицы невозможно, опираясь на официальные источники информации.

Однако, геоданные вряд ли заменят традиционные исследования по изучению мобильности людей, которые дают более глубокое понимание социально-экономических характеристик. Рациональное сочетание традиционных [4] и новых методов способно дать повышение качества, точности и оперативности знаний о локализации и мобильности населения.

Исследование выполнено за счет средств государственного задания (№ госрегистрации темы АААА-А21-121012190063-2) и при финансовой поддержке РФФИ в рамках научного проекта № 19-55-44020 Монг_т.

СПИСОК ЛИТЕРАТУРЫ

1. Бабкин Р.А. Пространственная динамика Московской агломерации // *Географические исследования Сибири и сопредельных территорий*, 2019, с. 22-25.
2. Воробьев А.Н. Картографирование плотности населения редкозаселенного региона (на примере Иркутской области) // *Геодезия и картография*, 2019, №4, с. 32-38.
3. Нырцов М.В., Нырцова Т.П. Большие данные в картографии. Умное картографирование: будущее или технологическое изменение // *Известия высших учебных заведений. Геодезия и аэрофотосъемка*, 2016, №5, с. 42-45.
4. Цветков В.Я., Домницкая Э.В. Геоданные как основа цифрового моделирования // *Современные наукоемкие технологии*, 2008, №4, с. 100-101.
5. Черняк Л. Свежий взгляд на Большие Данные // *Открытые системы СУБД*, 2017, №7, с. 48-51.

Конфликт интересов: Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Поступила в редакцию 28.05.2020

Принята к публикации: 28.05.2021

UDC 528.946 (911.3:314)

ISSN 1609-0683

DOI: <https://doi.org/10.17308/geo.2021.2/3447>

Use of Geodata for Operational Study of Population

A. N. Vorobyev ✉

*B. Sochava Institute of Geography SB RAS, Russian Federation
(1, Ulan-Batorskaya St., Irkutsk, 664033)*

Abstract: The purpose of the 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, the Russian and world experience in compiling maps based on mobile phones data was analysed.

Materials and methods. To achieve the goal, the author uses the data received from mobile operators (CDR), with the help of which it was possible to identify information about location of the device at each moment in time. Based on the geographical location of base stations, it is possible to analyze 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 certain time interval.

Results and discussion. The results obtained now clearly demonstrate that geographic data from user devices and geographic information systems represent a very effective symbiosis of technologies for solving geographic problems. Such studies make it possible to assess the general mobility of the population

© Vorobyev A. N., 2021

Alexander N. Vorobyev, e-mail: Tore12@yandex.ru



The content is available under Creative Commons Attribution 4.0 License.

for a particular region. Data on the location of a mobile phone can really reveal the spatio-temporal distribution of urban residents at macro- and mesolevels in the administrative districts of the city.

Conclusion. Based on official sources of information (statistical compilations and population censuses), it is impossible to quickly consider the movement of the population at the local level. However, for all its indisputable advantages, geodata are unlikely to replace traditional methods of studying human mobility in the near future. Only a rational combination of traditional and new methods can cause a qualitative change in the accuracy and efficiency of population distribution.

Key words: big data, geodata, mobile phone, population density, population mobility.

Funding: The work was carried out in the framework of the scientific work in the Institute of Geography named after V.B. Sochava of the Siberian Branch of the RAS (draft number 0347-2016-0004).

For citation: Vorobyev A.N. Use of Geodata for Operational Study of Population. *Vestnik Voronezskogo gosudarstvennogo universiteta. Seria: Geografia. Geoekologia*, 2021, no. 2, pp. 49-54. (In Russ.). DOI: <https://doi.org/10.17308/geo.2021.2/3447>

REFERENCES

1. Babkin R.A. Prostranstvennaya dinamika Moskovskoy aglomeratsii [Spatial dynamics of the Moscow agglomeration]. *Geograficheskie issledovaniya Sibiri i sosed'nykh territoriy*, 2019, pp. 22-25. (In Russ.)
2. Vorob'ev A.N. Kartografirovaniye plotnosti naseleniya redkozaselennogo regiona (na primere Irkutskoy oblasti) [Mapping of density of the population of a rarely populated region (on the example of the Irkutsk region)]. *Geodeziya i kartografiya*, 2019, no. 4, pp. 32-38. (In Russ.)
3. Nyrtsov M.V., Nyrtsova T.P. Bol'shie dannye v kartografii. Umnoe kartografirovaniye: budushchee ili tekhnologicheskoe izmeneniye [Big data in cartography. Smart mapping: Future or technological change]. *Izvestiya*

vysshikh uchebnykh zavedeniy. Geodeziya i aerofotos'emka, 2016, no. 5, pp. 42-45. (In Russ.)

4. Tsvetkov V.Ya., Domnitskaya E.V. Geodannye kak osnova tsifrovogo modelirovaniya [Geodata as the basis of digital modeling]. *Sovremennyye naukoemkie tekhnologii*, 2008, no. 4, pp. 100-101. (In Russ.)

5. Chernyak L. Svezhiy vzglyad na Bol'shie Dannye [Fresh look at big data]. *Otkrytye sistemy SUBD*, 2017, no. 7, pp. 48-51. (In Russ.)

Conflict of interests: The author declares no information of obvious and potential conflicts of interest related to the publication of this article.

Received: 28.05.2020

Accepted: 28.05.2021

Воробьев Александр Николаевич
кандидат географических наук, научный сотрудник лаборатории картографии, геоинформатики и дистанционных методов Института географии им. В.Б. Сочавы СО РАН, г. Иркутск, Российская Федерация, ORCID: 0000-0001-5529-1384, e-mail: Tore12@yandex.ru

Alexander N. Vorobyev
Cand. Sci. (Geogr.), Researcher of the Laboratory of Cartography, Geoinformatics and Remote Methods, V.B. Sochava Institute of Geography of the Siberian Branch of the RAS, Irkutsk, Russian Federation, ORCID: 0000-0001-5529-1384, e-mail: Tore12@yandex.ru