

УДК 314.728+331.556.2
JEL: O15, R23
DOI 10.24147/1812-3988.2021.19(1).58-69

АНАЛИЗ ДАННЫХ СОТОВЫХ СЕТЕЙ ПРИ ИЗУЧЕНИИ НАСЕЛЕНИЯ

Р.А. Долженко

Уральский институт управления – филиал РАНХиГС (Екатеринбург, Россия)

Информация о статье

Дата поступления
28 ноября 2020 г.

Дата принятия в печать
15 января 2021 г.

Тип статьи

Исследовательская статья

Ключевые слова

Геоаналитика, сотовые компании, мобильные операторы, большие данные, анализ данных, визуализация данных

Аннотация. Рассмотрены возможности использования больших данных, сосредоточенных у агрегаторов, обеспечивающих предоставление услуг мобильной связи и обладающих информацией о пользователях в разных аспектах их деятельности (перемещения, расходы, интересы, социально-демографические показатели) в рамках геоаналитики. Данное направление только начинает использоваться для принятия управленческих решений, что повышает актуальность предложенных в работе наработок. Показано, что использование анализа данных сотовых компаний позволяет оценить различные стороны жизнедеятельности населения в разбивке по любой территории, например миграционные потоки внутри региона и между регионами вплоть до персоны. Использование данных сотовых компаний позволяет обеспечить оперативность, точность анализа, как результат, повысить эффективность управленческих решений стейкхолдеров. Предложен набор данных, который необходим для подобной оценки на примере оператора мобильной связи. Описаны возможности, а также сложности, с которыми может столкнуться исследователь при анализе больших данных компаний, обеспечивающих услуги мобильной связи. Разработана методика анализа данных сотовых сетей, которая включает в себя различные параметры, необходимые для заказчика исследования. Приведены примеры базовых отчетов о населении районов с использованием данных мобильных сетей, которые могут быть использованы при анализе взаимозависимостей между параметрами для принятия управленческих решений в части развития миграционной политики в регионе. Предложенные наработки могут быть использованы в деятельности органов государственной и муниципальной власти, а также для повышения эффективности исследования демографических показателей на территориях страны посредством учета данных о конкретных субъектах.

DATA ANALYSIS OF MOBILE NETWORKS IN THE STUDY OF POPULATION

R.A. Dolzhenko

Ural Institute of Management – Branch of RANEPA (Yekaterinburg, Russia)

Article info

Received
November 28, 2020

Accepted
January 15, 2021

Type paper

Research paper

Keywords

Geoanalysis, cellular companies, mobile operators, Big Data, data analysis, data visualization

Abstract. The paper considers the possibilities of using big data, concentrated among aggregators that provide mobile services and have information about users in various aspects of their activities (movements, expenses, interests, socio-demographic indicators) in the framework of geoanalytics. This direction is just beginning to be used for making management decisions, which increases the relevance of the developments proposed in the work. It is shown that the use of data analysis of cellular companies makes it possible to assess various aspects of the life of the population broken down by any territory, for example, migration flows within a region and between regions up to a person. The use of data from cellular companies makes it possible to ensure the efficiency, accuracy of the analysis, as a result, to increase the effectiveness of management decisions of the stakeholders. A set of data is proposed, which is necessary for such an assessment on the example of a mobile operator. It describes the possibilities and difficulties that a researcher may face when analyzing the big data of companies that provide mobile services. A methodology for analyzing data from cellular networks has been developed, which includes various parameters required for the research customer. Examples of basic reports on the population of districts using data from mobile networks are given, which can be used to analyze the interdependencies between the parameters for making managerial decisions regarding the development of migration policy in the region. The proposed developments can be used in the activities of state and municipal authorities, as well as to increase the efficiency of the study of demographic indicators in the country's territories by considering data on specific subjects.

1. Введение. Ключевой тренд, который проявляет себя в современной экономике, – повсеместная цифровизация, под которой понимается не просто активное насыщение производственных и общественных процессов информационными системами, но переход к использованию данных для повышения их эффективности. И если данные активно аккумулируются многими участниками экономических отношений, особенно теми, кто приближен к своим клиентам, то направления их использования с помощью искусственного интеллекта, статистического и предиктивного анализа, машинного обучения, только начинают формироваться.

С другой стороны, происходящие изменения приводят к усилению миграционных потоков на территории Российской Федерации, которые уже в среднесрочной перспективе могут значительно изменить социально-трудовой ландшафт регионов страны. Для того чтобы управлять миграцией в новых условиях, необходимо использовать адекватные механизмы оценки ее потоков. Именно анализ больших данных о перемещениях людей на территориях может стать тем инструментом, который позволит в оперативном порядке получать срез ситуации, оперативно реагировать на изменения показателей, предсказывать их значение на перспективу.

Цель работы – провести оценку теоретико-методологических оснований для изучения населения с помощью анализа данных, собираемых компаниями сотовой связи, а также использования его результатов для принятия управленческих решений в различных областях общественной жизни. В статье категории «сотовая связь» и «мобильная связь», а также связанные понятия мы будем рассматривать в качестве синонимов, не вдаваясь в их различия, так как в контексте нашей работы они не существенны.

2. Обзор литературы. Обращаясь к теме использования больших данных для анализа миграционных потоков, прежде всего отметим, что работ, посвященных ей, в научных изданиях практически нет. Точнее, в настоящее время в русскоязычном сегменте научной литературы данное направление представлено лишь в одной работе в журнале «Вопросы статистики» [1]. В данной статье коллектив авторов рассматривает использование больших данных в целях получения новых знаний о населении и исследования на их основе раз-

личных социально-экономических явлений и процессов. Еще одна работа – исследователей под руководством В.И. Татаренко из Сибири – рассматривает возможности использования данных о населении для построения сети коммерческих объектов [2]. Ю.Ю. Соловьева рассматривает возможности использования различных методов анализа в геомаркетинге [3].

В зарубежных изданиях этих работ больше, но не на много. Контент-анализ публикаций, индексируемых в базах *Web of Science* и *Scopus*, показал, что по данной тематике за 2010–2019 гг. опубликовано чуть больше 20 работ, в которых изучаются перспективы исследования общественных процессов с помощью данных сотовых сетей, например в статьях F. Calabrese с соавторами [4], X. Zuo и Y. Zhang [5], в некоторых представлены конкретные направления использования, в частности: использование *GPRS* координат пользователей для определения потоков перемещения на примере трафика в Бостоне [6], оценка привлекательности инфраструктурных объектов с помощью распределения и плотности цифровых следов пользователей мобильных телефонов [7], изучение городского трафика с помощью социальных сетей в работах F. Calabrese с соавторами [8], S. Tao с соавторами [9], нагрузка на социальные сети как индикатор социальных событий на карте в статьях коллективов ученых под руководством J. Reades [10], R. Ahas [11], F. Girardin [12], обнаружение людей с помощью данных [13] и т. д. Отметим, что большая часть публикаций приходится на период с 2010 по 2012 г. И если первые публикации отражают возможности использования данных социальных сетей, то в дальнейшем исследователи описывают конкретные варианты применения различных математических моделей для анализа данных [14–16].

Отметим, что существует целый пласт исследований, которые затрагивают вопросы использования данных для геоаналитики или, если говорить о решении конкретных прикладных задач, геомаркетинга, как правило, они представлены в работах автора [17], С.Г. Кисельгоф, Ф.Т. Алескерова [18]. В основе данной технологии лежит использование данных при принятии решений по развитию регионов и городов с учетом миграционных потоков населения. Однако, именно в последние годы за счет бурного развития технологии машинного обучения и *Big Data* можно говорить о

прорывных решениях в данной области. Геоаналитика с помощью данных сотовых сетей выходит на качественно новый уровень. Но, прежде чем рассмотреть методические подходы к этому направлению, определимся с базовыми понятиями.

Можно дать следующее определение понятия «геоаналитика»: это инструмент для получения массивов данных о населении и его перемещениях, который построен на анализе

технологических событий агрегатора и построении математических, предиктивных моделей с использованием технологий *Big Data* и *Machine Learning*.

Повсеместная цифровизация процессов, активное подключение населения к Сети и генерация большого количества цифровых следов позволяет эффективно использовать собранные о жителях данные, что имеет ряд несомненных преимуществ использования (рис. 1).



Рис. 1. Преимущества использования технологии *Big Data* для анализа и управления территорией

Fig. 1. Advantages of using *Big Data* technology for analysis and management of the territory

Ключевая особенность данных, собираемых мобильными сетями, заключается в том, что они отражают реальную картину в оперативном режиме по всем пользователям, а в силу распространения мобильной связи можно утверждать, что данные репрезентативны для всего населения.

С другой стороны, нельзя обойти стороной и сложности, которые могут ухудшить результаты традиционных подходов к геоаналитике населения, к ним мы предлагаем ряд важных моментов. Во-первых, любые территории характеризуются несбалансированностью объектов транспортной, социальной, деловой инфраструктуры – каждый населенный пункт уникален, обладает особенностями, которые не позволяют делать общие выводы и рекомендации. Во-вторых, любой сбор данных о таких сложных объектах, как население крупных территорий, сопряжен с высокими расходами. В-третьих, анализ основывается на обобщениях и допущениях, которые могут приводить к иска-

жению выводов. Для достоверных результатов исследования необходима качественная проработка характера использования территории, ее истории, особенностей экономического и досугового ландшафта.

С нашей точки зрения, все эти ограничения могут быть сняты с помощью использования больших данных, генерируемых мобильными сетями о населении. В этом случае можно получить следующие результаты применения данных геоаналитики:

- использование показателей о численности и плотности населения с высокой пространственной детализацией с глубиной аналитики до уровня конкретного пользователя;
- наличие данных о траекториях потоков населения, в том числе очищенных от случайных перемещений;
- прогноз спроса на существующие и новые маршруты с учетом трендов, скрытой динамики изменения;

- в перспективе, когда стоимость данных будет снижена, можно спрогнозировать уменьшение уровня расходов заказчиков на транспортное моделирование.

Кроме того, использование *Big Data* позволяет получать актуальную информацию о дневной и вечерней корреспонденции населения, строить и анализировать трафики города, находить оптимальные маршруты и точки притяжения людей. Как видно, сфера применения данной технологии для управления территории очень разнообразна и, главное, эффективна, однако на текущем уровне принятия она слишком сложна для органов власти (в их штатном расписании просто отсутствуют специалисты с компетенциями в данной области), дорогá (приобретение массива данных у любого мобильного оператора стоит значительные для бюджета средства).

В любом случае, использование данных сотовых сетей позволит повысить эффективность управления городскими процессами, территориальным развитием, мобильностью населения, транспортными потоками и др. В современных условиях муниципалитет уже не может проводить культурно-массовые мероприятия, развивать туризм, планировать мероприятия по обеспечению безопасности, разрабатывать стратегию развития территории, строительство новых объектов и др. без использования данных сотовых сетей, а также их глубокого анализа, построенного на основе искусственного интеллекта и машинного обучения.

Рассмотрим, какие риски и возможности влечет за собой использование геоаналитики в различных сферах общественной жизни, например безопасности населения, туризма, градостроительства (конкретный выбор определен необходимостью отразить абсолютно разные направления деятельности).

3. Направления использования геоаналитики в различных сферах общественной жизни. Обеспечение безопасности граждан крайне важно в современных условиях, так как существуют значительные риски нарушения безопасности и угрозы для населения, кроме того, мы до сих пор не научились прогнозировать развитие ситуаций, анализировать состояние трендов, предсказывать опасные события, наконец, реакция на подобные риски должна быть очень оперативной, особенно по отношению к категориям населения, которые находятся в зоне риска. Применение данных геоанали-

тики в этой области позволит добиться высокой точности и оперативности мониторинга групп лиц, находящихся в зоне риска, определения мест их концентрации, маршрутов следования, мест прибытия и убытия. Также станет возможным прогнозирование развития рисков общественной безопасности, связанных с миграцией населения, с учетом накопленных за предыдущие периоды данных высокой точности и достоверности.

Туризм как сфера деятельности актуален для многих регионов страны. В силу ее больших размеров, расширение туристической отрасли затруднено из-за недостаточного и неравномерного развития необходимой инфраструктуры, недостаточной диверсификации предложений для различных сегментов, до сих пор отсутствует опыт сбора и анализа информации о въезжающих туристах, как результат – бизнес и органы власти имеют данные низкого качества о структуре и параметрах туристического потока и эффекте высокобюджетных мероприятий, направленных на его привлечение. В свою очередь, туризм является одной из самых перспективных сфер развития геоаналитики, которая позволит использовать актуальные данные о турпотоках, сократить временные и бюджетные затраты на сбор информации, начать использовать ранее недоступные данные, повысить эффективность планирования развития туризма и продвижения отдельных мероприятий, расширить перечень категорий туристов.

Использование больших данных в градостроительстве позволяет более эффективно управлять территориальным развитием. Как известно, чем крупнее населенный пункт, тем сложнее сбалансировать социальную, деловую инфраструктуру, их использование в разное время или отдельными категориями населения, как результат – у жителей может снижаться удовлетворенность качеством и комфортом среды.

Посредством учета численности населения и его потоков можно повысить удовлетворенность жителей через оптимальное размещение объектов социальной, транспортной, деловой инфраструктуры. Следствием увеличения масштаба сбора данных может стать значительная экономия на издержках сбора информации.

Подобные преимущества использования геоаналитики можно выделить практически для любых сфер общественной жизни. Ее вне-

дрение, как показано выше, ограничивается дороговизной данных и отсутствием необходимых компетенций в указанной области у органов власти, муниципалитетов, возможных заказчиков. На стоимость услуги может повлиять только конкуренция, а вот второе ограничение может быть преодолено за счет обучения сотрудников и набора необходимых специалистов на рынке. В любом случае заказчик должен обладать сформированным видением подходов к использованию подобных данных для решения своих задач. Именно их мы и представим далее.

4. Возможности использования данных сотовых сетей о населении. С опорой на работу авторов под руководством В.И. Татаренко [2], с нашей точки зрения анализ данных сотовых сетей в самом упрощенном виде включает в себя три этапа:

- сбор и подготовка данных;
- анализ данных;
- визуализация результатов.

Если говорить о более детальной декомпозиции задач, стоящих перед исследователями, то методика анализа будет иметь следующий вид (рис. 2).



Рис. 2. Методика анализа данных сотовых сетей

Fig. 2. Methods for analyzing data from cellular networks

Задача сбора и анализа данных является непрерывным процессом, который должен подстраиваться под запросы стейкхолдеров (заказчиков) оценки.

Круг базовых отчетов, результатов анализа данных о населении, которые могут использоваться в целях развития градостроительства и управления территориями, достаточно широк. Например, можно анализировать плотность населения, динамику его численности, распределение по районам и др. Базовое применение

данных мобильных сетей не многим отличается от статистики, которая использует для анализа изменения показателей социологические опросы и косвенные показатели.

Сотовые компании могут продавать как данные, так и результат их анализа. Считаем, что собственная аналитика будет более результативна за счет экономии на издержках, учета особенностей заказчика и его целей. Поэтому необходимо четко понимать, какие данные и в каком формате должны быть представлены

для анализа. Классический вариант – это табличная форма предоставления, структурированная по ряду базовых переменных.

Каким образом могут быть представлены данные о населении, его плотности, миграционных потоках? С учетом того, какие данные и в каком формате агрегируются сотовыми ком-

паниями, мы предлагаем следующую базовую структуру пакетов данных.

Самый простой пакет данных для анализа – «Изменение численности населения», он описывает количество людей, проживающих или работающих на заданной территории. Пример его представления приведен в табл. 1.

Таблица 1. Пример списка переменных пакета данных «Изменение численности населения»

Table 1. An example of a list of variables in the data package "Population change"

<i>Переменная</i>	<i>Описание</i>	<i>Формат</i>
Dt	Дата окончания периода, за который собирались данные	YYYY.MM.DD
Zid	Идентификатор зоны	Целое число
customers_cnt_home	Число человек, проживающих на заданной территории	Целое неотрицательное число
customers_cnt_job	Число человек, работающих на заданной территории	Целое неотрицательное число
customers_cnt_day	Число человек, которые обычно находятся на заданной территории днем	Целое неотрицательное число
customers_cnt_move	Число человек, проживающих на заданной территории и имеющих разъездную работу	Целое неотрицательное число

Второй базовый пакет данных для первичного анализа – «Изменение плотности населения», он описывает количество людей, нахо-

дящихся на заданной территории, с детализацией до получасовых интервалов (см. пример в табл. 2).

Таблица 2. Пример списка переменных пакета данных «Изменение плотности населения»

Table 2. An example of a list of variables of the data package "Population density change"

<i>Переменная</i>	<i>Описание</i>	<i>Формат</i>
Ts	Дата и время начала получасового интервала	YYYY.MM.DD HH:MM
Zid	Идентификатор зоны	Целое число
customers_cnt_total	Общее количество абонентов, находящихся на территории района (с учетом тех, кто находится на территории района непродолжительное время)	Целое неотрицательное число
customers_cnt_long	Число абонентов, находящихся на территории района продолжительное время	Целое неотрицательное число
customers_cnt_work	Число абонентов, находящихся на территории района на рабочем месте	Целое неотрицательное число
customers_cnt_loc	Число абонентов, находящихся на территории района у себя дома	Целое неотрицательное число
customers_cnt_city	Число абонентов, находящихся на территории района и проживающих в городе	Целое неотрицательное число
customers_cnt_territory	Число абонентов, находящихся на территории района и проживающих в регионе	Целое неотрицательное число

Последний базовый пакет, который необходим для общего представления о населении территории, – «Динамика перемещений», он описывает количество людей, совершивших

поездку между любыми сочетаниями локаций, с детализацией до получасовых интервалов (см. пример в табл. 3).

Таблица 3. Пример списка переменных пакета данных «Динамика перемещений»
 Table 3. An example of the list of variables of the data package "Displacement dynamics"

Переменная	Описание	Формат
Ts	Дата и время начала интервала, за который предоставляется матрица	YYYY.MM.DD HH:MM
departure_zid	Идентификатор зоны отправления	Целое число
arrival_zid	Идентификатор зоны прибытия	Целое число
customers_cnt	Число человек, начавших поездку в заданный временной интервал	Целое неотрицательное число
customers_cnt_metro	Число человек, совершивших поездку на метро	Целое неотрицательное число
customers_cnt_home_work	Число человек, совершивших поездку из дома на работу	Целое неотрицательное число
customers_cnt_work_home	Число человек, совершивших поездку с работы домой	Целое неотрицательное число
customers_cnt_returned	Число человек, совершивших поездку и вернувшихся обратно	Целое неотрицательное число
customers_cnt_end	Число человек, закончивших поездку в заданный временной интервал	Целое неотрицательное число
customers_cnt_static	Число неподвижных человек	Целое неотрицательное число

Перемещение из своей зоны и возврат в нее (в течение указанного интервала) фиксируются в диагонали матрицы ($arrival_zid = departure_zid$). Отсутствие движения абонента фиксируется нулевой зоной назначения ($arrival_zid = 0$).

В таких базовых пакетах находится необходимый запас данных для проведения практически любой аналитики. Она ограничена используемыми сегментами, а также инструментами анализа: фильтрацией данных, перекрестным анализом, машинным обучением, *Look-*

alike (подходом к отбору схожей аудитории) и др. По отзывам экспертов, которые занимаются аналитикой, можно выделить около 400 различных сегментов: семейный статус, состояние трудоустройства, наличие автомобиля и характер его использования, туризм, наличие льгот, использование социальных, государственных и прочих сервисов, градация по уровню доходов, статус образования, отношение к спорту, наличие домашних животных и т. п. Пример базовых сегментов приведен в табл. 4–6.

Таблица 4. Примеры сегментации данных мобильных сетей о населении по социально-демографическому портрету

Table 4. Examples of segmentation of mobile networks data on the population by socio-demographic portrait

Параметр	Сегменты
Пол	Мужчины
	Женщины
Возраст	18–24
	25–34
	35–44
	45–54
	55–63
	64+
Уровень дохода, тыс. руб. в месяц	0–25
	25–75
	75–150
	150+
Занятость	Учащиеся
	Работающие
	Предприниматели

Окончание табл. 4
The end of Table 4

<i>Параметр</i>	<i>Сегменты</i>
Наличие детей	0–16
	0–3
	4–9
	10–16

Таблица 5. Примеры сегментации данных мобильных сетей о населении по геопоказателям

Table 5. Examples of segmentation of mobile networks data on the population by geo-indicators

<i>Параметр</i>	<i>Территория, на которой доступен параметр</i>
Координаты / адрес работы	Москва и Московская область
Координаты / адрес дома	Москва и Московская область
Координаты / адрес утреннего пребывания (топ-5 локаций)	Россия
Координаты / адрес вечернего пребывания (топ-5 локаций)	Россия
Регион проживания	Россия
Путешествующие за границу	Россия
Путешествующие за границу в развитые страны	Россия
Путешествующие по России	Россия
Известны регионы / города и частота поездок	Россия
Известны страны и частота поездок за рубеж	Россия
Готовые совершить поездку за границу	Россия

Таблица 6. Примеры сегментации данных мобильных сетей о населении по прочим критериям

Table 6. Examples of segmentation of data of mobile networks about the population by other criteria

<i>Параметр</i>	<i>Территория, на которой доступен параметр</i>
Автомобилисты	Москва и Московская область
Беременные женщины	Москва
Надежные клиенты для выдачи кредита	Россия
Заинтересованные в кредите + надежные клиенты	Россия
Не гражданин Российской Федерации	Россия
Недвижимость	Москва и Московская область
День рождения в ближайшие 7 / 14 / 30 дней	Россия
Покупки в Интернете	Россия
Владельцы смартфонов	Россия
Криптовалютчики	Россия
Фотографы	Москва и Московская область
Заинтересованные в потребительском кредите	Россия
Семьи	Россия
Заинтересованные в спутниковом телевидении	Россия
Заинтересованные в изучении английского языка	Москва и Московская область
Слушают музыку	Москва и Московская область
Читают книги (со смартфонов, <i>e-book</i>)	Москва и Московская область
Ездят на метро (1–3 раза в месяц, более 4 раз)	Москва и Московская область
Имеют девайсы дороже 50 000	Россия
Заинтересованные в ипотеке	Москва и Московская область
Шопоголики (масс-маркет и премиум)	Москва и Московская область
Водители такси	Москва и Московская область
Покупают через смартфон (<i>m-commerce</i>)	Москва и Московская область
В поиске работы	Москва и Московская область
Домохозяйки	Москва и Московская область

Также возможны прочие сегменты по тематике потребляемого контента: мобильная связь, банковские услуги, мультимедиа, реклама, социальные сети, такси, госуслуги, косметика, образование и т. д.

После того как сегменты определены и зафиксированы переменные, которые нас интересуют, проводится выборка данных. На ее основе проводится анализ с использованием наиболее адекватных инструментов.

Так, для понимания транзитного потока населения по территории нами была построена регрессионная модель:

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n,$$

где Y – зависимая переменная, x – влияющий фактор, a – коэффициент.

В зависимости от кластера формула регрессионной модели может приобретать следующий вид:

$$\begin{aligned} \text{Транзитный поток} = & \text{базовая метрика} + \\ & + \text{миграции} + \text{путешествия} + \text{семьи} + \\ & + \text{период времени} \end{aligned}$$

или

$$\begin{aligned} \text{Загрузка метрополитена} = & \text{базовая метрика} + \\ & + \text{пол} + \text{возраст} + \text{уровень дохода} + \\ & + \text{занятость} + \text{дети}. \end{aligned}$$

Для анализа данных могут использоваться штатный функционал *Microsoft Excel*, *Stata*, *SPPS*, *R*, которые имеют широкий функционал для обработки данных, в том числе с наложением на карты.

После того как проведен анализ данных, важно также правильно представить его результаты. Часто этот момент оказывается более важным для принятия управленческих решений, чем глубина анализа и сложность используемой математической модели. Опять же, упомянутые программные комплексы для анализа данных обладают интегрированными графическими инструментами с опорой на самые современные достижения в сфере визуализации. Отметим, что очень хороший функционал графического представления данных имеется в программах *Power BI* и *Tableau*.

На рис. 3 приведен пример графического отображения данных о количестве людей, находящихся на Красной площади во время проведения масштабного мероприятия.

Среднее в интервале с 10:00 до 22:00 число иностранцев и жителей регионов России, одновременно* находившихся на Красной площади (по датам работы Парка Футбола)

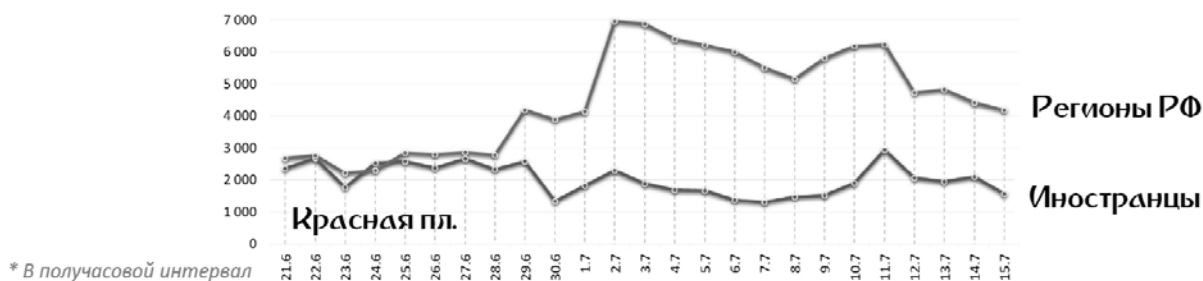


Рис. 3. Пример графического отображения данных о количестве людей на территории

Fig. 3. An example of graphical display of data on the number of people in the territory

Данные могут быть также представлены в формате тепловой карты как по базовым станциям, так и по квадратам размером 500×500 м (в силу ограничения технологии сотовой связи). Для построения карт территориальной доступности можно использовать пакет для визуализации пространственных данных *ggmap* с использованием онлайн-источника *Google Maps* или *API* Яндекс.Карт. Географические координаты строений могут быть получены с помощью *Google Geolocation API*. Графическая интерпретация данных может быть осу-

ществлена с помощью набора компонентов *GIS Toolkit Active*.

Для обоснования насыщенности потоков населения могут быть использованы данные о территории (численность, уровень безработицы, наличие образовательных организаций, транспортные остановки, количество рабочих мест на территории и т. д.) из базы *Ruslana*.

Таков не полный перечень инструментов и подходов к анализу данных мобильных операторов и использованию результатов в управлении городской инфраструктурой для пони-

мания миграционных потоков и принятия стратегических решений, влияющих на население территории. Мы не будем претендовать на глубину анализа и представления возможностей, которые открываются перед исследователями населения, когда они начнут использовать большие данные мобильных операторов, так как это слишком сложная и многогранная задача. Но первые шаги в этом направлении уже делаются, поэтому важно понимать, что может получить общество сейчас, какие перспективы будут потом.

5. Заключение. Проведенный нами обзор методических оснований исследования данных о населении с помощью данных сотовых сетей о пользователях показал, что у данного направления статистического анализа большие перспективы. Работа с действительно большим,

чистым массивом данных, возможность использования инструментов машинного обучения и *Big Data*, эффективная визуализация результатов анализа данных и многие другие преимущества позволяют сделать вывод о перспективах данной области. С другой стороны, у возможных заказчиков и пользователей этого вида данных недостаточно компетенций не только для изучения, но и для понимания целесообразности подобных исследований. Монополизация данных (сейчас в стране три крупнейших компании, оказывающих услуги сотовой связи) приводит к тому, что цены на них действительно завышены, но уже в среднесрочной перспективе можно предположить, что они упадут. И значит, уже сейчас нужно формировать и развивать компетенции исследователей в этом направлении.

Литература

1. Полунина М. В., Ельникова Е. А., Аветисян С. Т. Новый цифровой источник статистической информации о населении // Вопросы статистики. – 2018. – № 25 (1). – С. 74–85.
2. Татаренко В. И., Вдовин С. А. Основные этапы плана геомаркетингового и геоаналитического исследования коммерческих объектов // Инновации и инвестиции. – 2019. – № 4. – С. 119–123.
3. Соловьева Ю. Ю. Использование методов стратегического анализа и прогноза в геомаркетинге // Интерэкспо Гео/Сибирь. – 2017. – Т. 6. – № 2. – С. 77–80.
4. Calabrese F., Colonna M., Lovisolo P., Parata D., Ratti C. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome // IEEE Transactions on Intelligent Transportation Systems. – 2011. – Vol. 12 (1). – P. 141–151.
5. Zuo X., Zhang Y. Detection and Analysis of Urban Area Hotspots Based on Cell Phone Traffic // Journal of Computers. – 2012. – Vol. 7 (7). – P. 1753–1760.
6. Diao M., Zhu Y., Ferreira J., Ratti C. Inferring Individual Daily Activities from Mobile Phone Traces: A Boston Example // Environment and Planning B: Planning and Design. – 2016. – Vol. 43, iss. 5. – P. 920–940. – DOI: 10.1177/0265813515600896.
7. Sevtsuk A., Ratti C. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks // Journal of Urban Technology. – 2010. – Vol. 17 (1). – P. 41–60.
8. Calabrese F., Ferrari L., Blondel V. D. Urban Sensing Using Mobile Phone Network Data: A Survey of Research // ACM Computing Surveys. – 2014. – Vol. 47 (2). – P. 1–23.
9. Tao S., Vasileios M., Rodriguez S., Rusu A. Real-Time Urban Traffic State Estimation with A-GPS Mobile Phones as Probes // Journal of Transportation Technologies. – 2012. – No. 2. – P. 22–31.
10. Reades J., Calabrese F., Sevtsuk A., Ratti C. Cellular census: Explorations in urban data collection // IEEE Pervasive Computing. – 2007. – Vol. 6. – No. 3. – P. 30–38.
11. Ahas R., Silm S., Jarv O., Saluveer E., Tiru M. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones // Journal of Urban Technology. – 2010. – Vol. 17 (1). – P. 3–27.
12. Girardin F., Vaccari A., Gerber A., Biderman A., Ratti C. Quantifying urban attractiveness from the distribution and density of digital footprints // International Journal of Spatial Data Infrastructures Research. – 2009. – Vol. 4. – P. 175–200.
13. Gonzalez M. C., Hidalgo C. A., Barabasi A. L. Understanding individual human mobility patterns // Nature. – 2008. – Vol. 453. – P. 479–482.
14. Terada M., Nagata T., Kobayashi M. Population Estimation Technology for Mobile Spatial Statistics // NTT DOCOMO Technical Journal. – 2013. – Vol. 14. – No. 3. – P. 10–15.

15. Toole J. L., Ulm M., Bauer D., González M. C., Bauer D. Inferring land use from mobile phone activity // *UrbComp'12 : Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. – New York, NY, USA : Association for Computing Machinery, 2012. – P. 1–8. – DOI: 10.1145/2346496.2346498.
16. Traag V. A., Browet A., Calabrese F., Morlot F. Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Inference // *IEEE Third International Conference on Social Computing*. – Boston, MA, USA, 2011. – DOI: 10.1109/PASSAT/SocialCom.2011.133.
17. Долженко Р. А. Геомаркетинговое моделирование расположения сети офисов продаж банка // *Менеджмент в России и за рубежом*. – 2016. – № 1. – С. 41–50.
18. Кисельгоф С. Г., Алескеров Ф. Т. Размещение отделений банка. Обзор проблемы // *Бизнес-информатика*. – 2009. – № 1 (07). – С. 59–69.

References

1. Polunina M.V., El'nikova E.A., Avetisyan S.T. Novyi tsifrovoy istochnik statisticheskoi informatsii o naselenii. *Voprosy statistiki*, 2018, no. 25 (1), pp. 74-85. (in Russian).
2. Tatarenko V.I., Vdovin S.A. Osnovnye etapy plana geomarketingovogo i geoanaliticheskogo issledovaniya kommercheskikh ob'ektov. *Innovatsii i investitsii*, 2019, no. 4, pp. 119-123. (in Russian).
3. Solov'eva Yu.Yu. Ispol'zovanie metodov strategicheskogo analiza i prognoza v geomarketinge. *Interesko Geo'Sibir'*, 2017, Vol. 6, no. 2, pp. 77-80. (in Russian).
4. Calabrese F., Colonna M., Lovisolo P., Parata D., Ratti C. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems*, 2011, Vol. 12 (1), pp. 141-151.
5. Zuo X., Zhang Y. Detection and Analysis of Urban Area Hotspots Based on Cell Phone Traffic. *Journal of Computers*, 2012, Vol. 7 (7), pp. 1753-1760.
6. Diao M., Zhu Y., Ferreira J., Ratti C. Inferring Individual Daily Activities from Mobile Phone Traces: A Boston Example. *Environment and Planning B: Planning and Design*, 2016, Vol. 43, iss. 5, pp. 920-940. DOI: 10.1177/0265813515600896.
7. Sevtsuk A., Ratti C. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology*, 2010, Vol. 17 (1), pp. 41-60.
8. Calabrese F., Ferrari L., Blondel V.D. Urban Sensing Using Mobile Phone Network Data: A Survey of Research. *ACM Computing Surveys*, 2014, Vol. 47 (2), pp. 1-23.
9. Tao S., Vasileios M., Rodriguez S., Rusu A. Real-Time Urban Traffic State Estimation with A-GPS Mobile Phones as Probes. *Journal of Transportation Technologies*, 2012, No. 2, pp. 22-31.
10. Reades J., Calabrese F., Sevtsuk A., Ratti C. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 2007, Vol. 6, no. 3, pp. 30-38.
11. Ahas R., Silm S., Jarv O., Saluveer E., Tiru M. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology*, 2010, Vol. 17 (1), pp. 3-27.
12. Girardin F., Vaccari A., Gerber A., Biderman A., Ratti C. Quantifying urban attractiveness from the distribution and density of digital footprints. *International Journal of Spatial Data Infrastructures Research*, 2009, Vol. 4, pp. 175-200.
13. Gonzalez M.C., Hidalgo C.A., Barabasi A.L. Understanding individual human mobility patterns. *Nature*, 2008, Vol. 453, pp. 479-482.
14. Terada M., Nagata T., Kobayashi M. Population Estimation Technology for Mobile Spatial Statistics. *NTT DOCOMO Technical Journal*, 2013, Vol. 14, no. 3, pp. 10-15.
15. Toole J.L., Ulm M., Bauer D., González M.C., Bauer D. Inferring land use from mobile phone activity, in: *UrbComp'12*, Proceedings of the ACM SIGKDD International Workshop on Urban Computing, New York, NY, USA, Association for Computing Machinery publ., 2012, pp. 1-8. DOI: 10.1145/2346496.2346498.
16. Traag V.A., Browet A., Calabrese F., Morlot F. Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Inference, in: *IEEE Third International Conference on Social Computing*, Boston, MA, USA, 2011. DOI: 10.1109/PASSAT/SocialCom.2011.133.

17. Dolzhenko R.A. Geomarketingovoe modelirovanie raspolozheniya seti ofisov prodazh banka. *Menedzhment v Rossii i za rubezhom*, 2016, no. 1, pp. 41-50. (in Russian).

18. Kisel'gof S.G., Aleskerov F.T. Razmeshchenie otdelenii banka. Obzor problemy. *Biznes-informatika*, 2009, no. 1 (07), pp. 59-69. (in Russian).

Сведения об авторе

Долженко Руслан Алексеевич – д-р экон. наук, доцент, директор

Адрес для корреспонденции: 620144, Россия, Екатеринбург, ул. 8 Марта, 66

E-mail: snurk17@gmail.com

ORCID: 0000-0003-3524-3005

Scopus AuthorID: 56436594700

ResearcherID: J-2847-2015

РИНЦ AuthorID: 8576-4140

About the author

Ruslan A. Dolzhenko – Doctor of Economic Sciences, Director

Postal address: 66, 8 Marta ul., Yekaterinburg, 620144, Russia

E-mail: snurk17@gmail.com

ORCID: 0000-0003-3524-3005

Scopus AuthorID: 56436594700

ResearcherID: J-2847-2015

RSCI AuthorID: 8576-4140

Для цитирования

Долженко Р. А. Анализ данных сотовых сетей при изучении населения // Вестник Омского университета. Серия «Экономика». – 2021. – Т. 19, № 1. – С. 58–69. – DOI: 10.24147/1812-3988.2021.19(1).58-69.

For citations

Dolzhenko R.A. Data analysis of mobile networks in the study of population. *Herald of Omsk University. Series "Economics"*, 2021, Vol. 19, no. 1, pp. 58-69. DOI: 10.24147/1812-3988.2021.19(1).58-69. (in Russian).