

Новый цифровой источник статистической информации о населении

Мария Владимировна Полунина,
Елена Анатольевна Ельникова,
Сергей Тарханович Аветисян
ООО «ИнфоНет Мобил», г. Москва, Россия

В статье дается изложение авторского понимания вопросов использования больших данных в целях получения новых знаний о населении и исследовании на их основе различных социально-экономических явлений и процессов. Указывается, что с развитием информационной индустрии и повсеместным проникновением мобильной связи, как показывает зарубежный опыт, одним из наиболее перспективных источников больших данных с точки зрения полноты охвата населения как объекта статистического наблюдения и оперативности получения информации о нем являются данные сотовых операторов. Авторы подробно рассматривают отечественный опыт в этой области. В частности, еще в 2014 г. в России удалось реализовать проект «Геоаналитика» по практическому применению данных операторов в управленческой деятельности органов исполнительной власти Правительства Москвы.

Авторами освещается история возникновения и развития нового цифрового источника статистической информации о населении, в основе которого лежат технические данные сотовых сетей. Излагаются основы инновационных российских методологических разработок и алгоритмов преобразования радиочастотных событий базовых станций операторов сотовой связи в статистические показатели численности, плотности и динамики перемещений населения с полным охватом исследуемой территории и высоким уровнем пространственно-временной детализации. Особое внимание в статье уделяется вопросам защиты персональных данных абонентов, легальности процессов сбора и обработки информации, получаемой от операторов сотовой связи, в соответствии с требованиями законодательства Российской Федерации.

Детально рассматриваются направления практического применения статистических показателей, сформированных на основе данных сотовых операторов, в области экономики, торговли, культуры, транспортного моделирования, сферах городского планирования и управления.

Ключевые слова: большие данные, цифровые источники информации, данные сотовых операторов, проект «Геоаналитика», статистическая информация о населении, геоаналитические показатели.

JEL: C10, C80.

Для цитирования: Полунина М.В., Ельникова Е.А., Аветисян С.Т. Новый цифровой источник статистической информации о населении. Вопросы статистики. 2018;25(1):74-85.

New Digital Source of Statistical Information about Population

Maria V. Polunina,
Elena A. Elnikova,
Sergey T. Avetisyan
InfoNet Mobil LLC, Moscow, Russia

The article presents the authors' views on using «Big Data» to gain new information on population and to study various social and economic phenomena and processes on its basis. As the foreign experience clearly demonstrates, with the development of information industry and the ubiquitous mobile communications penetration, one of the most promising sources of «Big Data» in terms of population coverage (using population as an object of statistical observation) and efficiency in obtaining information on it, is data from mobile operators. The paper also notes Russian experience in this field, especially since 2014, when Russia managed to implement the «Geoanalysis» project on using data from mobile operators in managerial activities of the Moscow Government.

The authors outline the history and development of the new digital data source for population statistics, which is based on technical data of cellular networks. The paper covers baselines to Russian innovation methodological developments and algorithms for converting radio frequency events from base stations of mobile operators into statistical indicators of number, density and dynamics of population movements with full coverage of study area and high level of space-time specification. This article pays particular attention to issues concerning protection of subscribers' personal data, legality of collecting and processing the information received from mobile operators in accordance with the legislation of the Russian Federation.

The authors thoroughly examined directions for applying statistical indicators, based on data from mobile operators, in the field of economy, trade, culture, transport modeling, urban planning and management.

Keywords: Big Data, digital sources of information, data from mobile operators, «Geoanalysis» project, statistical information on population, indicators of geoanalysis.

JEL: C10, C80.

For citation: Polunina M.V., Elnikova E.A., Avetisyan S.T. New Digital Source of Statistical Information about Population. *Voprosy statistiki*. 2018; 25(1):74-85. (In Russ.)

Статистическая информация о населении, предоставляемая органами официальной статистики, играет важную роль в государственном управлении. Наиболее полным источником данных о численности и структуре населения, его национальном и языковом составе, социально-экономических характеристиках являются переписи населения, проводимые каждые 10 лет.

Однако в настоящее время под влиянием научно-технического прогресса и информатизации развитие городской среды идет чрезвычайно быстрыми темпами, в связи с чем возникла потребность в получении дополнительной информации о населении для решения определенного круга задач в сфере городского планирования и управления.

В последние годы на международных научных конференциях все чаще поднимается вопрос о возможностях использования новых цифровых источников информации (больших данных) в целях получения дополнительных знаний о населении и исследовании на их основе различных социально-экономических явлений и процессов.

Сегодня, в связи со стремительным развитием информационной индустрии и повсеместным проникновением мобильной связи, одним из наиболее перспективных источников больших данных, с точки зрения полноты охвата населения, являются данные сотовых операторов. Изучение потенциала их практического использования и перспективы интеграции с традиционными источниками официальной статистики являются современным направлением как теоретических, так и прикладных исследований.

Анализ научной литературы показывает актуальность и перспективность использования данных сотовых операторов для изучения городской среды и поведенческих моделей населения [1-5]. Имеющиеся научные разработки очерчивают широкий круг задач, при решении которых подобные данные могут быть задействованы, а также описывают ряд методов, которые могут лежать в основе получения требуемых показателей [6-13].

Данные сотовых операторов могут использоваться для получения на их основе статистической информации о:

- численности населения и его распределении по территориям фактического проживания и работы;
- динамике посещаемости различных территорий города;
- взаимосвязях частей города между собой по ряду параметров;
- потоках маятниковой трудовой миграции;
- характеристиках транспортных потоков;
- транспортной доступности территорий города;
- прочее.

Тем не менее полностью готовых алгоритмов, которые напрямую могут быть использованы для получения на основе данных сотовых операторов информации о численности проживающего и работающего населения, наиболее посещаемых территориях города по времени суток, характеристиках транспортных потоков и пр., в научной литературе не представлено. Многие вопросы описываются поверхностно или не рассматриваются совсем.

Основные причины такого положения дел следующие:

- отсутствие у большинства исследователей достаточных компетенций в области телеком-индустрии, необходимых для понимания того, как организована работа сотовых сетей, какие данные и с каким уровнем точности могут быть получены;
- большая сложность в организации взаимодействия с операторами мобильной связи по изучению данных, с соблюдением законодательства в области защиты персональных данных;
- необходимость обрабатывать чрезвычайно большие объемы данных, что требует наличия высокого уровня информационно-вычислительной инфраструктуры.

В России были решены все вышеописанные вопросы и реализован полноценный проект «Геоаналитика», который уже несколько лет составляет на регулярной основе статистическую

информацию о численности, плотности, динамике перемещений населения, получаемую на основе данных сотовых операторов, органам исполнительной власти Правительства Москвы для практического использования в управленческой деятельности.

В публикуемой статье описан разработанный в рамках проекта «Геоаналитика» новый цифровой источник статистической информации о населении, в основе которого лежат данные операторов мобильной связи, и рассмотрены основные направления его практического применения органами исполнительной власти.

Структура работы выглядит следующим образом. В первой части приведен краткий обзор предпосылок возникновения и развития в России проекта «Геоаналитика». Во второй части подробно рассмотрен вопрос легальности получения данных от сотовых операторов в целях формирования статистических показателей о населении («геоаналитических» данных). В третьей части изложены методические основы формирования таких показателей. В заключительной части статьи описан опыт и направления практического применения «геоаналитических» данных в различных сферах городского планирования и управления.

1. Предпосылки возникновения и развития проекта «Геоаналитика». Первой предпосылкой возникновения в 2013 г. проекта «Геоаналитика» стало развитие и повсеместное проникновение сотовой связи. Активный рост абонентской базы в России завершился к 2010 г., достигнув при этом значений, сильно опережающих аналогичные показатели в Европейском союзе и странах СНГ. Так, уровень проникновения услуг сотовой связи в Российской Федерации (количество SIM-карт на 100 жителей) в 2010 г. составил 166,3% против 122-128% в Европе и СНГ. При этом доля россиян, имеющих хотя бы одну SIM-карту, по разным оценкам, выросла до 85-90%. Вместе с этим более чем в три раза (относительно 2000 г.) увеличился и средний объем потребляемых абонентом минут - с 90 до 300 минут к 2010 г.¹

В связи со значительным увеличением спроса на услуги сотовой связи в базах сотовых операторов стало аккумулироваться огромное количество

информации, наибольшую долю в общем объеме которой составляют данные о нагрузке на сотовые сети (радиочастотные сигнальные события), собираемые операторами в целях оперативного реагирования на возможные сбои в работе их обслуживания. Так, объем хранимых данных на серверах только одного сотового оператора несколько лет назад уже мог достигать 1200 терабайт в год.

Экспоненциальный рост информации, аккумулируемой в базах сотовых операторов, привел к возникновению научно-исследовательского направления по изучению потенциала прикладного использования данных сотовых операторов и получения на их основе статистической информации о населении. Однако широко используемые на тот момент технологии последовательной обработки данных не позволяли оперативно обрабатывать массивы информации такого объема. Активное развитие с 2011 г. технологий массивно-параллельной высокоскоростной обработки огромных массивов данных (BigData) стало второй предпосылкой создания в России проекта «Геоаналитика».

Наконец, третьей предпосылкой послужило появление на рынке квалифицированных BigData специалистов, которые умеют работать с большими данными.

Указанные предпосылки, а также высокая заинтересованность Правительства Москвы в получении дополнительных статистических данных о населении стали ключевыми факторами создания в 2013 г. проекта «Геоаналитика».

На первом этапе реализации проекта была проведена научно-исследовательская работа по изучению технических возможностей операторов мобильной связи и формированию методик получения из технических данных сотовых сетей (радиочастотных событий базовых станций) статистических показателей численности, плотности и динамики перемещений населения по территории Московского региона².

В рамках научно-исследовательской работы были достигнуты следующие практические результаты:

1. Определен и формализован отраслевой перечень статистических показателей о населении, являющихся значимыми для Комитета по архитектуре и градостроительству города Москвы в его работе по планировке территории столицы;

¹ Текст пресс-релиза оператора сотовой связи «Билайн» в официальном блоге на сайте Habrahabr от 31 мая 2012 г. URL: <https://habrahabr.ru/company/beeline/blog/144944/>.

² Здесь и далее под Московским регионом подразумевается территория, включающая в себя два субъекта РФ - Москву и Московскую область.

2. Разработаны методики сбора и обработки технических данных сотовых сетей в целях получения статистических показателей о населении, с учетом технических возможностей оборудования операторов и легальности выполняемых операций на всех этапах реализации методик;

3. Созданы методики объединения информации от всех трех федеральных операторов и последующей конвертации этих данных в статистические показатели о населении (в соответствии с отраслевыми требованиями).

На основе результатов научно-исследовательской работы был разработан набор геоаналитических отчетов, содержащих в общей сложности 52 статистических показателя о населении, которые Комитет по архитектуре и градостроительству города Москвы в 2014 г. задействовал в своей работе по градостроительному проектированию территорий столицы. О положительных результатах работ было доложено мэру и руководителям органов исполнительной власти города Москвы, что привело к появлению спроса на данные «Геоаналитики» со стороны различных департаментов Правительства Москвы, необходимости формирования геоаналитических отчетов на постоянной основе и послужило стимулом для дальнейшего развития проекта.

Рост спроса на данные «Геоаналитики», а также возникновение потребности в разработке дополнительных статистических показателей о населении в целях использования их при решении различных бизнес-задач, стоящих перед органами исполнительной власти, вызвали необходимость

разработки унифицированного словаря терминов, операционализирующих такие понятия, как проживающее население, работающее население и пр. Создание единой терминологии, основанной на описании поведенческих моделей населения, и ее использование в различных сегментах городского планирования позволили существенно сократить время и расходы на формирование дополнительных показателей.

На текущий момент разработано уже более 100 геоаналитических показателей, на основе которых можно исследовать численность, плотность и динамику перемещений населения на больших территориях в различных срезах и с высоким уровнем пространственно-временной детализации. Основные индикаторы, отображающие развитие проекта «Геоаналитика» с 2014 по 2017 г., представлены на рис. 1.

Проект «Геоаналитика» получил высокую оценку на уровне аппарата мэра и Правительства Москвы, ряд департаментов которого использует сегодня геоаналитические данные в своей работе на постоянной основе. За время существования проекта получено большое количество положительных отзывов от органов исполнительной власти города Москвы о целесообразности практического использования геоаналитических данных в области городского планирования и управления. В последнее время к проекту проявлен высокий интерес также со стороны других субъектов РФ. В частности, геоаналитические показатели были рассчитаны для территории г. Казани в целях использования их при разработке Генерального плана города.



Рис. 1. Увеличение спроса на данные «Геоаналитики» и их объема с 2014 по 2017 г.

*Департаментов Правительства Москвы и подведомственных им организаций.

2. Легальность получения данных от сотовых операторов. Наиболее частый вопрос, который возникает у специалистов и экспертов из различных отраслей при первичном ознакомлении с геоаналитическими показателями, формируемыми на основе данных, предоставляемых сотовыми операторами, касается легальности получения таких данных в соответствии с положениями Федерального закона «О персональных данных».

Для подтверждения соблюдения требований законодательства Российской Федерации в области персональных данных в процессе формирования геоаналитических показателей ниже подробно описывается состав данных, получаемых от сотовых операторов, и алгоритмы их обработки.

В базах операторов мобильной связи содержится огромное количество данных, которые можно разделить на две категории:

1. Персональные данные абонентов, которые включают в себя информацию из контрактной базы (ФИО, паспортные данные и т. д.) и биллинговые данные о ежемесячных расходах на мобильную связь, используемых абонентом тарифных планах, подключенных услугах и пр.;

2. Технические данные, которые генерируются при взаимодействии мобильных устройств с сотовой сетью и собираются операторами для мониторинга работы оборудования.

На всех этапах реализации методик, разработанных в рамках проекта «Геоаналитика», собираются и обрабатываются только технические данные (радиочастотные сигнальные события базовых станций), включающие в себя события следующих видов:

- события, связанные с использованием мобильного устройства абонентом, - звонки, SMS-сообщения, включение/выключение устройства (при сборе и обработке информации содержание голосовых/текстовых сообщений не задействуется, учитывается лишь сам факт звонка/отправки или принятия SMS-сообщения);

- события, связанные с переходом абонентского устройства в другую географическую зону (переключением устройства с одной группы базовых станций, объединенных по территориальному признаку, на другую);

- служебные сетевые события, которые происходят независимо от активности пользователей и включают в себя периодические события обновления местоположения абонентского устройства.

Весь процесс - от сбора данных сотовыми операторами до получения геоаналитических показателей - делится на два больших этапа.

На *первом этапе*, который выполняется сотовыми операторами на собственном оборудовании по разработанным в проекте «Геоаналитика» методикам, производятся сбор и обработка технических данных сотовых сетей, в результате которой формируются агрегированные наборы количественных данных, содержащие информацию, необходимую для последующего получения на ее основе геоаналитических показателей.

При сборе и обработке исходных данных операторами используется ряд математических алгоритмов, разработанных в рамках проекта:

- алгоритмы извлечения информации по заданным параметрам из общей базы данных;

- алгоритмы фильтрации извлеченных данных по ряду критериев. Одним из таких критериев являются объемы голосового и SMS-трафика, генерируемого абонентскими устройствами. Алгоритмы фильтрации радиочастотных событий на основе этого показателя позволяют исключить из объема обрабатываемых данных, собираемых с сетевого оборудования, трафик, создаваемый дополнительными SIM-картами в мульти-SIM устройствах, дополнительными телефонами (если абонент пользуется несколькими телефонами), а также планшетами, смарт-часами и любыми другими устройствами передачи данных, работающих по стандартам GSM;

- алгоритмы перевода информации о местоположении фиксируемых событий в координаты на исследуемой территории;

- алгоритмы по увеличению точности данных, так как события генерируются абонентскими устройствами не регулярно, а только при наступлении определенных условий (во время входящих/исходящих вызовов, отправки/принятия SMS-сообщений, при переключении устройства между базовыми станциями и пр.);

- алгоритмы формирования агрегированных показателей численности абонентских устройств с их распределением по территории во времени, которые (показатели) на последующих этапах конвертируются в статистические данные о населении. Показатели, в зависимости от типа их пространственно-временной детализации и смысловых характеристик, komponуются в различные наборы данных (отчеты), на основании которых невозможно определить местонахождение/действия/перемещения отдельно взятых абонентов.

На *втором этапе* сформированные на стороне сотовых операторов агрегированные наборы данных передаются в проект «Геоаналитика», где верифицируются, консолидируются и затем конвертируются в геоаналитические отчеты, содержащие статистические показатели о населении. Более подробно данный этап будет рассмотрен в следующем разделе.

Вышеприведенные положения методик разработки геоаналитических показателей обосновывают легальность получения для их формирования данных от сотовых операторов в соответствии с требованиями Федерального закона от 27.07.2006 № 152-ФЗ «О персональных данных», что официально подтверждено соответствующим заключением Роскомнадзора.

3. Основы формирования геоаналитических показателей. Системы сбора данных, телекоммуникационное оборудование и конфигурация сетей сотовых операторов различаются по ряду технических характеристик, что накладывает ограничения на возможности детализации данных, предоставляемых операторами для проекта «Геоаналитика». Для получения качественных данных от всех федеральных операторов методики формирования геоаналитических показателей разрабатывались с учетом технических особенностей организации сотовой сети каждого из них. При этом были определены единые критерии максимально возможной пространственно-временной детализации информации и периодичности ее предоставления.

Сегодня максимально возможный уровень деления территории Московского региона на расчетные зоны представляет собой сетку, разбитую на секторы 500×500 м. Общее количество секторов, на которые разделена территория региона, составляет 190242 единицы. Часть такого разбиения, для примера, представлена на рис. 2.

Такой высокий уровень пространственного деления территории позволяет получать более детальную информацию о местах проживания и работы населения, динамике изменения численности населения на территории в течение суток и т. д. для решения определенного круга бизнес-задач, стоящих перед органами исполнительной власти. Наиболее часто используемым типом деления территории Московского региона на расчетные зоны, по которым рассчитываются все геоаналитические показатели, является деление по административным районам.

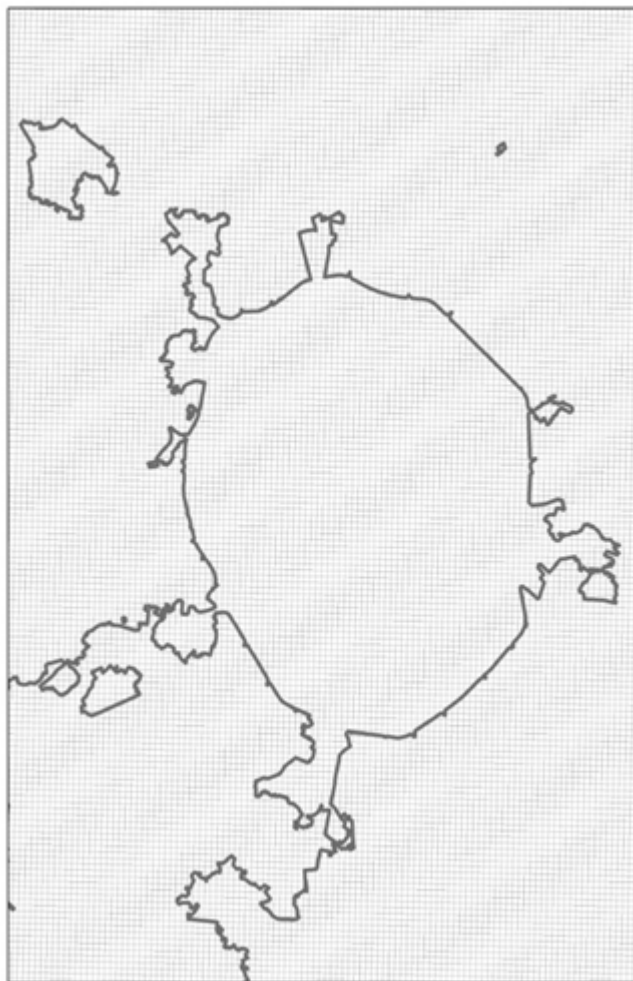


Рис. 2. Разбиение части территории Московского региона на секторы 500×500 м

Максимальный уровень временной гранулярности данных составляет 30 минут. Ряд геоаналитических показателей рассчитывается по получасовым интервалам каждого дня месяца. Такая точность данных необходима для исследования транспортных потоков, изучения динамики посещаемости населением различных территорий в течение суток и пр.

Исходный объем обрабатываемых данных, собираемых с сетевого оборудования, в среднем составляет 100 Тб в месяц и требует значительных программно-аппаратных мощностей. В связи с этим процесс формирования итоговых агрегированных наборов данных, необходимых для расчета геоаналитических показателей, сегодня занимает 2–3 недели. Однако развитие технологий высокоскоростной обработки данных позволяет говорить о сокращении в ближайшем будущем этих временных затрат, в которые входит не только сам процесс формирования на стороне

операторов агрегированных наборов данных, но и верификация их результатов, которая выполняется уже в проекте «Геоаналитика».

Методики верификации наборов данных, получаемых от сотовых операторов, являются важной частью процесса формирования геоаналитических показателей, так как позволяют выявлять «выбросы» в данных, связанные с техническими сбоями в работе базовых станций или возникающие в процессе сбора и/или обработки исходной информации на стороне сотовых операторов.

Методики верификации включают в себя алгоритмы проверки по критериям следующих видов:

- проверки на внутреннюю непротиворечивость значений взаимосвязанных показателей, содержащихся в наборе данных (например, численность проживающих на территории не может быть меньше, чем численность проживающих на этой территории и имеющих разъездную работу, так как второе значение всегда является подмножеством первого);

- проверки на непротиворечивость значений каждого показателя во времени (например, количество поездок с использованием метро по будням не должно сильно варьироваться между рабочими днями недели);

- перекрестные проверки значений взаимосвязанных показателей из разных наборов данных каждого оператора (например, на территории у себя дома одновременно не может находиться больше абонентов, чем на этой территории проживает в соответствующем месяце);

- встречные проверки между наборами данных, получаемых от разных операторов (например, средняя продолжительность поездок между двумя районами города по данным каждого из операторов должна быть примерно одинаковой);

- соотнесение с внешними открытыми источниками информации (например, проверяется, связан ли резкий всплеск численности абонентов, одновременно находящихся на территории, с техническими сбоями в работе базовых станций или же обоснован проведением какого-либо массового мероприятия, например празднования Дня Победы).

В зависимости от верифицируемого набора данных количество критериев проверки его качества варьируется в среднем от 15 до 35 штук.

При выявлении несоответствия наборов данных каким-либо из вышеперечисленных критериев сотовые операторы итерационно дорабатывают отчеты (наборы данных), пока их

качество не будет удовлетворять всем критериям проверок.

На заключительной стадии, выполняемой в проекте «Геоаналитика», верифицированные наборы данных консолидируются и конвертируются в геоаналитические отчеты, содержащие статистические показатели о населении. Для конвертации используются уникальные, разработанные в рамках проекта «Геоаналитика» математические модели, в основе которых лежат масштабирующие поправочные коэффициенты, учитывающие меняющийся уровень проникновения сотовых операторов на территории, а также взаимосвязи различных показателей.

Важно отметить, что статистические показатели о населении, формируемые в рамках проекта «Геоаналитика», являются дополнением данных официальной статистики, так как несут в себе иную смысловую нагрузку. Так, например, если по методологии Росстата численность проживающего на территории населения определяется на основе ответов респондентов об их постоянном (обычном) месте жительства, то по методике «Геоаналитики» показатель «численность проживающих на территории» отражает информацию о местах регулярного ночного длительного пребывания населения и его численности в этих локациях ежемесячно.

Таким образом, с одной стороны, схожие статистические показатели официальной статистики и «Геоаналитики» отражают различные стороны одного и того же явления/события, что затрудняет их сопоставление для оценки качества геоаналитических данных, с другой - их совместное использование позволяет решать более широкий круг различных прикладных задач. Подтверждением того, что геоаналитические данные валидны, является высокая оценка их качества самими заказчиками, которые используют эти данные в своей работе на постоянной основе, а также новые запросы на поставку геоаналитических отчетов и/или разработку дополнительных статистических показателей о населении.

Примерами геоаналитических отчетов, формируемых сегодня на ежемесячной основе, являются следующие:

- отчет «Изменение численности населения», отображающий ежемесячное распределение населения по территориям проживания/работы;

- отчет «Изменение плотности населения», отображающий динамику изменения численности проживающего/работающего/транзитного

населения на территории в течение суток каждого дня месяца;

- отчет «Матрица корреспонденций дом-работа»³, отображающий взаимосвязь территорий проживания и работы;

- отчет «Матрица корреспонденций перемещений населения между всеми районами города»⁴ с детализацией информации до получасовых интервалов каждого дня месяца;

- отчет «Продолжительность поездок», содержащий информацию как о завершённых, так и о транзитных поездках по территории города, с распределением их по длительности с точностью до 10 минут;

- отчет «Загрузка метрополитена»;

- другие.

Эти геоаналитические отчеты используются сегодня при решении широкого круга прикладных задач в различных сферах городского планирования и управления, транспортного моделирования, области экономики, торговли, культуры и пр.

4. Направления практического применения данных «Геоаналитики». Ключевыми эффектами от использования геоаналитических данных в своей работе органами исполнительной власти являются следующие:

- экономия финансовых затрат и времени на сбор и обработку информации, необходимой для принятия обоснованных управленческих решений в области социально-экономического планирования;

- возможность увеличить поступления в городской бюджет за счет более эффективного планирования различных мероприятий в области экономики;

- дополнительные эффекты, в зависимости от области применения данных «Геоаналитики».

Так, например, применение геоаналитических показателей в градостроительном планировании позволяет повысить качество комплексного проектирования территорий за счет:

- использования статистических показателей с высокой пространственной детализацией;

- уменьшения погрешностей в расчетах за счет высокой точности данных;

- получения актуальных данных на ежемесячной основе, позволяющих учитывать сезонные

факторы и отслеживать тренды в изменении численности населения на территориях;

- возможности мониторить изменения в поведенческих моделях населения, связанные с проведением тех или иных градостроительных мероприятий.

Наиболее крупными проектами, для использования в которых были предоставлены геоаналитические данные о населении в различных срезах, являются комплексное проектирование развивающихся территорий Новой Москвы, а также разработка Генерального плана столицы.

В социальной сфере, помимо планирования градостроительных мероприятий, данные «Геоаналитики» могут использоваться и для определения уровня востребованности и характера посещаемости населением территорий расположения социальных объектов и объектов культуры. Например, геоаналитические данные о численности проживающего/работающего/транзитного населения, находящегося на территории города, разбитой по сетке 500 × 500 м, и с детализацией до получасовых интервалов, а также матрица корреспонденций перемещений населения между районами города были предоставлены двум департаментам Правительства Москвы для решения ими следующих задач:

- Департаменту культуры города Москвы требовались данные для изучения динамики посещаемости территорий, примыкающих к местам расположения московских библиотек, в целях корректировки и оптимизации графиков их работы;

- Департаменту городского имущества города Москвы данные нужны были для изучения поведенческого портрета аудитории ВДНХ в целях разработки коммуникативной стратегии по увеличению уровня посещаемости данного объекта культуры.

Сбалансированность социальной и транспортной инфраструктур является необходимым условием успешного социально-экономического развития города. Инструментом планирования и оценки развития дорожно-транспортной инфраструктуры являются математические транспортные модели, разработкой которых занимаются подведомственные организации Департамента транспорта и развития дорожно-транспортной инфраструктуры города Москвы.

³ В отчете содержится информация по каждой паре районов региона о численности населения, проживающего в одном из них и при этом работающего в другом.

⁴ В отчете содержится информация по каждой паре районов региона о численности перемещений населения между ними (с детализацией до получасовых интервалов каждого дня месяца).

Для построения транспортных моделей используется информация из различных источников: данные об организации улично-дорожной сети и маршрутах общественного транспорта, данные из автоматизированных систем сбора (датчики движения транспорта, турникеты в метрополитене и т. д.), а также статистическая информация о населении.

Основная ценность геоаналитических данных применительно к транспортному моделированию заключается в том, что они содержат информацию о численности перемещений населения между всеми районами города и области с детализацией до получасовых интервалов каждого дня месяца. Кроме того, ежемесячное обновление данных «Геоаналитики» о количестве проживающего и работающего населения по сетке 500×500 м, охватывающей всю территорию агломерации, позволяет существенно повысить точность математических транспортных моделей, рассчитываемых в масштабах всего города.

В настоящее время Департамент транспорта и развития дорожно-транспортной инфраструктуры города Москвы использует геоаналитические данные в своей работе на постоянной основе.

Примерами практических задач по транспортному моделированию, реализованных непосредственно в рамках проекта «Геоаналитика», являются следующие:

- разработка мульти-агентной транспортной модели столицы в целях определения объема потенциального спроса на проектируемую систему из девяти линий скоростного автобусного транспорта Москвы и расчета оптимальных показателей для запуска первого из маршрутов;

- оценка изменения спроса на междугородные автобусные маршруты Московского региона за 2014-2016 гг. и визуализация результатов на интерактивной ВІ-платформе в целях проведения анализа данных в различных срезах.

В области экономики данные «Геоаналитики» могут использоваться для изучения на их основе различных показателей, связанных с численностью работающего населения, а также потоков маятниковой трудовой миграции между всеми районами города и области.

Геоаналитические данные содержат ряд показателей, на основе которых можно определить:

- ежемесячную численность работающего на территории населения (с детализацией до секторов 500×500 м);

- соотношение численности городского и областного населения, работающего в каждом районе региона;

- районы, в которые на работу приезжает наибольшее количество населения (относительно других районов города/области);

- отношение численности населения, проживающего и работающего в районе, ко всему проживающему в данном районе населению, имеющему работу;

- уровень замещаемости рабочих мест (на основе соотношения численности населения, проживающего в районе, но работающего за его пределами, к численности населения, работающего в этом районе, но проживающего за его пределами);

- численность поездок «дом-работа» и «работа-дом» с детализацией до получасовых интервалов каждого дня месяца между всеми районами города и области;

- прочее.

Объединение данных «Геоаналитики» с собственными данными заказчиков позволяет решать более широкий круг задач на основе разработки сложных математических моделей в целях выявления/предсказания латентных характеристик объектов/процессов.

Так, например, для Департамента экономической политики и развития города Москвы была разработана методика выявления на территории столицы потенциальных торговых и офисных объектов, которые уклоняются от уплаты налога на имущество. В основе методики лежала математическая модель, построенная на объединенных данных «Геоаналитики» и информации от департамента о жилой и нежилой недвижимости, объектах, которые платят налог, и тех, которые освобождены от его уплаты.

Потребность в разработке такой методики заключалась в следующем. Для выявления торговых и офисных объектов, обязанных платить налог на имущество, в Москве используются ежегодные обходы инспекторов, список объектов проверки для которых формируется на основе официальных данных. Поскольку такой информации недостаточно, это ведет к не очень высокому проценту обнаруживаемых релевантных объектов из общего списка обхода.

На основе разработанной методики был сформирован приоритезированный список потенциальных торговых и офисных объектов, состоящий

из 4000 позиций⁵, который был предоставлен для его включения в работу инспекторов, что, в свою очередь, создало возможность увеличения количества и скорости обнаружения торговых и офисных объектов, уклоняющихся от уплаты налога на имущество.

Данные «Геоаналитики» могут использоваться и при решении задач в области развития потребительского рынка. Так, Департамент торговли и услуг города Москвы регулярно проводит специализированные ярмарки и фестивали цикла «Московские сезоны». Анализ посещаемости ярмарок с помощью геоаналитических данных позволяет определить различные критерии эффективности их проведения и выявить:

- наиболее посещаемые ярмарки в разрезе будних и выходных дней;
- уровень прироста населения на территории расположения ярмарок в период их проведения;
- средние и максимальные значения численности населения, находившегося на территории проведения ярмарок в часы их работы (с детализацией до получасовых интервалов);
- дни и время, когда на территории ярмарок одновременно находилось наибольшее количество посетителей;
- среднюю численность посещений ярмарок, в том числе иногородним населением, учитывающим в своем составе туристов;
- уровень востребованности ярмарок среди городского/областного населения.

На основе вышеприведенной информации, дополненной данными о локациях и часах наибольшей концентрации населения на территории города (полученных также на основе анализа геоаналитических показателей), можно корректировать места и время проведения ярмарок/фестивалей, тем самым повышая их посещаемость.

Потенциал данных «Геоаналитики» позволяет говорить о возможности прикладного использования их и при решении задач, стоящих перед другими департаментами Правительства Москвы.

Также хотелось бы отметить, что кроме исполнительных органов власти, геоаналитические данные предоставляются сегодня в научно-исследовательских целях университетам, например

МГУ имени М.В. Ломоносова и НИУ «Высшая школа экономики».

Заключение. В настоящее время потенциал практического использования больших данных, и в частности данных сотовых операторов, изучается во многих странах. Так, например, в Великобритании в 2016 г. было проведено экспериментальное исследование по определению мест проживания и работы населения Лондона на основе анализа данных одного из сотовых операторов. Для изучения использовалась обезличенная информация о местоположении абонентов в ночное и дневное время. Результаты исследования показали высокий уровень соответствия данным последней переписи населения. В рамках проведенной научной работы изучался также вопрос использования данных сотовых операторов для определения этнической принадлежности, уровня доходов и численности населения⁶.

Подобные исследования по расчету численности населения на основе данных сотовых операторов проводились чуть ранее в Бельгии и Италии, причем в обоих случаях была отмечена тесная корреляция с данными административного учета.

В России же результаты научно-исследовательских работ в этой области и сотрудничество со всеми федеральными сотовыми операторами позволили уже в 2014 г. реализовать проект «Геоаналитика» по практическому применению данных операторов в управленческой деятельности органов исполнительной власти города Москвы.

Основным прорывом в области использования данных сотовых операторов стало то, что для формирования статистических показателей о населении (геоаналитических данных) задействуются только технические данные сотовых сетей (радиочастотные события базовых станций); персональные данные абонентов полностью исключаются из процесса сбора и обработки информации. Высокой точности геоаналитических данных позволило добиться объединение информации сразу от всех сотовых операторов, обслуживающих территорию Московского региона.

Сегодня данные «Геоаналитики» используются департаментами Правительства Москвы при

⁵ Для понимания того, насколько процесс разработки методики и формирования на ее основе приоритизированного списка был трудоемкий, надо отметить, что общий объем обработанных данных для получения результирующего набора в 4 тыс. строк составил несколько миллиардов строк, включающих в себя геоаналитические данные и информацию о недвижимости, предоставленную Департаментом экономической политики и развития города Москвы.

⁶ ONS Research Outputs: Using mobile phone data to estimate commuting flows. URL: <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics/researchoutputsusingmobilephonedatatoestimatecommutingflows>.

решении широкого круга практических задач в области экономики, торговли, культуры, транспортного моделирования, сферах городского планирования и управления.

Развитие технологических возможностей сотовых операторов позволяет говорить об увеличении в ближайшем будущем максимального уровня разбиения территории на расчетные зоны с сетки 500×500 м до секторов 30×30 м. Это открывает новые возможности в области исследований применимости геоаналитических данных при решении других прикладных задач в вышеперечисленных сферах (например, по выявлению зданий с наибольшим количеством торговых площадей или обнаружению апарт-ментов в нежилых зданиях - для Департамента экономической политики и развития г. Москвы).

Схожесть характера и ритмов жизни крупных городов позволяет оперативно внедрять и применять на новых территориях уже доказавшие свою эффективность методики формирования геоаналитических показателей. Так, например, всего за несколько месяцев данные методики были адаптированы для территории г. Казани, на основе которых были рассчитаны статистические показатели о населении в целях использования их при разработке Генерального плана города.

В настоящее время потенциал использования данных «Геоаналитики» видится в расширении круга задач, решаемых с их помощью органами исполнительной власти города Москвы, налаживании сотрудничества с другими субъектами РФ по поставке геоаналитических данных, а также их интеграции в систему официальной статистики.

Литература

1. 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.
2. Diao M., Zhu Y., Ferreira J., Ratti C. Inferring Individual Daily Activities from Mobile Phone Traces: A Boston Example // IRES Working Paper Series. 2015. IRES2015-012.
3. Gonzalez M.C., Hidalgo C.A., Barabasi A.L. Understanding individual human mobility patterns // Nature. 2008. Vol. 453. P. 479-482.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
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. 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.
11. Toole J.L., Ulm M., Bauer D., Gonzalez M.C. Inferring land use from mobile phone activity // Proceedings of the ACM SIGKDD International Workshop on Urban Computing. 2012. P. 1-8.
12. 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.
13. 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.

Информация об авторах

Полунина Мария Владимировна - руководитель отдела аналитики, ООО «ИнфоНет Мобил». 127051, г. Москва, ул. Неглинная, д. 17, стр. 2. E-mail: M.polunina@in-data.io.

Ельникова Елена Анатольевна - операционный директор, ООО «ИнфоНет Мобил». 127051, г. Москва, ул. Неглинная, д. 17, стр. 2. E-mail: E.elnikova@in-data.io.

Аветисян Сергей Тарханович - генеральный директор, ООО «ИнфоНет Мобил». 127051, г. Москва, ул. Неглинная, д. 17, стр. 2. E-mail: Avetisyanst@in-data.io.

References

1. 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;12(1):141-151.
2. Diao M., Zhu Y., Ferreira J., Ratti C. *Inferring Individual Daily Activities from Mobile Phone Traces: A Boston Example*. IRES Working Paper Series. 2015. IRES2015-012.
3. Gonzalez M.C., Hidalgo C.A., Barabasi A.L. Understanding Individual Human Mobility Patterns. *Nature*. 2008;453:479-482.
4. Reades J., Calabrese F., Sevtsuk A., Ratti C. Cellular Census: Explorations in Urban Data Collection. *IEEE Pervasive Computing*. 2007;6(3):30-38.
5. Sevtsuk A., Ratti C. Does Urban Mobility Have a Daily Routine? Learning from the Aggregate Data of Mobile Networks. *Journal of Urban Technology*. 2010;17(1):41-60.
6. 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;17(1):3-27.
7. Calabrese F., Ferrari L., Blondel V.D. Urban Sensing Using Mobile Phone Network Data: A Survey of Research. *ACM Computing Surveys*. 2014; 47(2):1-23.
8. 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;4:175-200.
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;(2):22-31.
10. Terada M., Nagata T., Kobayashi M. Population Estimation Technology for Mobile Spatial Statistics. *NTT DOCOMO Technical Journal*. 2013;14(3):10-15.
11. Toole J.L., Ulm M., Bauer D., Gonzalez M.C. Inferring Land Use from Mobile Phone Activity. *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. 2012. P. 1-8.
12. 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.
13. Zuo X., Zhang Y. Detection and Analysis of Urban Area Hotspots Based on Cell Phone Traffic. *Journal of Computers*. 2012;7(7):1753-1760.

About the Authors

Maria V. Polunina - Head of the Analytics Department, InfoNet Mobil LLC. 17/2, Neglinnaya st., Moscow, Russia, 127051. E-mail: M.polunina@in-data.io.

Elena A. Elnikova - Chief Operating Officer, InfoNet Mobil LLC. 17/2, Neglinnaya st., Moscow, Russia, 127051. E-mail: E.elnikova@in-data.io.

Sergei T. Avetisyan - Chief Executive Officer, InfoNet Mobil LLC. 17/2, Neglinnaya st., Moscow, Russia, 127051. E-mail: Avetisyanst@in-data.io.