

Использование данных телекоммуникационных операторов в транспортном планировании

Д.Е. Намиот, О.Н. Покусаев, А.Е. Чекмарев

Аннотация— В статье рассматриваются вопросы, связанные с использованием данных телекоммуникационных операторов в транспортном планировании. В последнее время, проникновение мобильных телефонов обеспечило то, что именно данные, которые собирают телекоммуникационные операторы становятся основным инструментом измерения перемещений в городах. Можно без преувеличения сказать, что именно с таких данных и началась цифровая урбанистика. Речь идет о том, что телекоммуникационные операторы естественным образом (для собственных целей биллинга услуг связи) собирают информацию о присутствии мобильных устройств в разных районах обслуживания. Более точно – об обслуживании мобильного устройства различными базовыми станциями оператора, каждая из которых и привязана к некоторой географической области. Соответственно, на уровне оператора понятно, когда конкретное мобильное устройство переместилось из одной области в другую (перешло на обслуживание от одной базовой станции к другой). Такие обезличенные и агрегированные по времени данные дают информацию о количестве мобильных устройств (количестве владельцев этих устройств) которые перемещались из одного района в другой за заданный временной интервал. Например, за 15 минут, 30 минут, один час и т.д. Такое представление людских потоков и является базой для транспортного планирования.

Ключевые слова—цифровая урбанистика, транспортное планирование, моделирование, прогнозирование.

I. ВВЕДЕНИЕ

В работе мы хотим остановиться на использовании данных, собираемых телекоммуникационными операторами в транспортном моделировании (планировании). О чем идет речь, когда говорят о данных телекоммуникационных операторов? В процессе обслуживания телекоммуникационных устройств (мобильных телефонов и т.д.) операторы связи собирают большое количество служебной информации. В первую очередь, эти данные, естественно, необходимы для

поддержки собственных бизнес-процессов операторов. Именно для этого они и собираются. В частности, файлы (логи, журналы) с записями деталей разговоров (взаимодействия) – так называемые Call Details Records (также используется термин Call Data Records) [1] служит основой для биллинга. Иными словами – это база для оценки (расчета) экономики оператора. По сути – основной набор данных, необходимый оператору. Эти записи, естественно, могут использоваться и для других целей (например, поиск и проверка информации о местоположении конкретного абонента и т.п.), но первичной остается именно экономическая составляющая. А это, в свою очередь, означает, что такая информация становится сравнительно “дешевым” источником данных для сторонних организаций. Не нужно заниматься созданием собственных служб сбора данных, установкой и обслуживанием сенсоров и т.п., если научиться получать полезную информацию из данных операторов, которые оператор (операторы) собирают в любом случае и которые есть для любых мест (районов и т.д.), где есть абоненты этих операторов. А проникновение (распространение) мобильных устройств обеспечивает то, что автоматически сбор данных будет охватывать все регионы. Именно эти соображения и определяют интерес, который проявляется исследователями в отношении данных мобильных операторов.

Естественно, данные, которые собирают операторы не являются (не должны являться) единственными в урбанистических и транспортных исследованиях. Например, информация об использовании проездных документов (карт) пассажирами может непосредственно давать информацию о маршрутах [2]. В каких-то случаях информация о перемещениях может собираться специально установленными программами на мобильных телефонах. Вопрос, однако, состоит в том, что с данными операторов трудно конкурировать в смысле охвата. Естественно, что перемещения пассажиров (мобильных телефонов), которые перемещаются, например, в пригородной электричке будут отражаться и в данных операторов. Из данных, собираемых операторами, будет трудно выделить именно железнодорожных пассажиров, но зато эти данные будут содержать информацию и об остальных перемещениях, для которых нет никакой информации о

Статья получена 21 сентября 2019.

Д.Е. Намиот - МГУ имени М.В. Ломоносова (e-mail: dnamiot@gmail.com); РУТ(МИИТ).

О.Н. Покусаев - Центр цифровых высокоскоростных транспортных систем РУТ (МИИТ) (email: o.pokusaev@rut.digital)

А.Е. Чекмарев - Центр цифровых высокоскоростных транспортных систем РУТ (МИИТ) (email: a.chekmarev@rut.digital)

валидациях проездных документов [3].

Что именно собирают операторы? Самое главное, что здесь нужно отметить – это то, что в данном случае речь идет об информации, относящейся к активности их абонентов. По сути – сам факт какой-то активности (звонок, сообщение и т.д.), а также место (с какой-то точностью, естественно), где это произошло. Здесь речь не идет о записи содержания звонков, например, или о чем-то подобном. Это – отдельные задачи, которые не имеют отношения к анализу данных в цифровой урбанистике или транспортном моделировании.

CDR – это типичные метаданные. В запись включается информация о номере, осуществившем звонок, о номере, принявшем звонок, длительность разговора, идентификация устройства и т.д. [4]. Местоположение определяется с использованием информации о базовых станциях, обслуживающих мобильное устройство в каждый конкретный момент времени [5].

Технически у оператора, естественно, существует возможность связать данные из CDR с профайлом пользователя и, соответственно, использовать еще и информацию из профайла (пол, возраст и т.д.). Но это может не делаться из соображений приватности. Также, из соображений приватности, такого рода связи вряд ли будут доступны сторонним организациям, занимающимся анализом данных. Поэтому такого данные далее не рассматриваются, оценка возраста и пола возможна и без данных профайлов (по схожести поведения). Также из соображений приватности идентификация мобильных устройств (если она вообще присутствует) обычно заменяется односторонним хэшем. Еще необходимо отметить, что в данных CDR, предоставляемых для анализа (опять-таки, из соображений приватности) могут быть изменены временные метки для событий, специально “загрублены” координаты и т.д. [6].

Дальнейшая часть статьи структурирована следующим образом. В разделе II кратко рассматривается история вопроса. Раздел III посвящен собственно методам анализа данных.

II. ИСПОЛЬЗОВАНИЕ “ТЕЛЕФОННЫХ” ДАННЫХ

В настоящем разделе мы хотели бы кратко изложить историю данного вопроса. Это важно для понимания того, откуда взялись (выросли) такого рода исследования, что находилось в их основе. Второй момент – это то, что многие изначально существовавшие подходы и идеи оказались, на наш взгляд, как-то забыты. И это абсолютно незаслуженно.

По нашему мнению, первенство в этом вопросе принадлежит MIT с серией работ по социальной динамике. Собственно говоря, они и предложили этот термин: *social dynamics* [7]. Для того, чтобы анализировать поведение и взаимодействие людей было предложено автоматически собирать данные об их активности. Так возникло направление *reality mining* [8].

Технически, это можно описать как тепловые карты, которые строились на основе данных о пользовательской активности. Иными словами – где и что (с точки зрения операций, доступных на телефоне) делали пользователи. При этом важно отметить, что сбор данных зависел от приложения, установленного на телефоне пользователя. Очевидно, что такой подход будет не очень масштабируемым (как убедить произвольных пользователей ставить служебное приложение, которое, по сути, записывает их действия?). Хотя для каких-то ограниченных целей это вполне может работать [9]. Наиболее простой пример – служебные телефоны, режим использования которых определяется работодателем [10]. Также, довольно большая часть предлагаемых корпорациями мобильных приложений занимается также (чаще всего – без явного объявления) сбором данных об активности пользователей, что разработчики надеются, в последующем, использовать в коммерческих целях.

Собственно говоря, именно идея масштабирования и привела к необходимости использования данных, которые автоматически собираются телекоммуникационными операторами. Приоритет здесь также принадлежит MIT, а первой практической работой на этом пути была, видимо, работа 2008 года *Real Time Rome* [11], а также последующая серия работ по городскому мониторингу [12]. Первое, что позволяла делать предложенная платформа – это отслеживать перемещения (или определять, что перемещения нет) для мобильных устройств. Также была введена возможность эмпирического разделения пешеходов и устройств, находящихся в транспорте. Система была ориентирована на мониторинг показателей в реальном времени, так что ее типичные отчеты следует рассматривать в привязке к некоторому моменту времени (временному интервалу):

- Где люди находятся в течение дня
- Концентрация и перемещение людей во время специальных событий
- Какие исторические места в городе привлекают наибольшее количество людей
- Перемещения пешеходов и телефонов (пользователей) находящихся в транспорте
- Перемещение и концентрация иностранцев (для Рима важно было отслеживать туристов). Они определялись по IMSI (*International Mobile Subscriber Identity*) номеру телефона [13], относящегося к мобильным сетям зарубежных операторов.

Интересно, что в этой системе, поддерживаемой *Telecom Italia*, собирались также данные с GPS-приемников, установленных в автобусах и такси. Соответственно, можно было сопоставлять транспортные данные с информацией о мобильных абонентах.

Очевидно, что последнее замечание об интеграции данных имеет еще больше оснований к реализации в 2018 году, когда общественный транспорт полностью оборудуется устройствами GPS. Попросту, совместное использование данных откроет возможность оценки загрузки пассажирского транспорта. Особенно интересным будет синергетический эффект, когда появится возможность анализировать GPS треки автомобилей. Как правило, именно подсчет автомобилистов представляет собой сложную, но очень актуальную, задачу городской аналитики.

В еще более обобщенном виде, последнее утверждение можно представить как необходимость использования операторских данных в сочетании с информацией из других источников. Иными словами, успех применения аналитики данных заключается в использовании так называемых мэшапов [14].

III. АНАЛИЗ “ТЕЛЕФОННЫХ” ДАННЫХ

В этом разделе мы хотели бы остановиться, собственно, на анализе данных.

Естественно, что большая часть исследований продолжила идею *reality mining* и использовала данные для выявления, например, социальных связей [15]. Данные, которые собирают операторы, касаются, естественно, и собственно “телефонной” активности. Где мобильные абоненты совершали звонки? Можно ли выделить какие-то шаблоны в событиях “позвонить по телефону”. Типичные задачи представлены, например, в статье [16]. Это, очевидно, имеет косвенное отношение к транспорту, поэтому далее это направление здесь не рассматривается.

Выше уже были рассмотрены задачи отображения перемещений, которые решались в проекте *Real Time Rome*. Естественно, что для поддержки этих перемещений нужна транспортная инфраструктура (дороги, общественный транспорт и т.д.). Соответственно, основное, что нужно определять – это шаблоны перемещений. Главный вопрос, для ответа на который нужны данные телекома – как люди перемещаются? Естественно, и это необходимо еще раз отметить, мы ведем речь именно о транспортных задачах.

Вот одна из наиболее цитируемых работ [17]. Основная задача – определение плотности населения по активности мобильных устройств, регистрируемой операторами. На рисунке 1 – картина для Португалии.

Математический аппарат – диаграммы Вороного [18]. Диаграмма Вороного конечного множества точек S на плоскости представляет такое разбиение плоскости, при котором каждая область этого разбиения образует множество точек, более близких к одному из элементов множества S , чем к любому другому элементу множества.

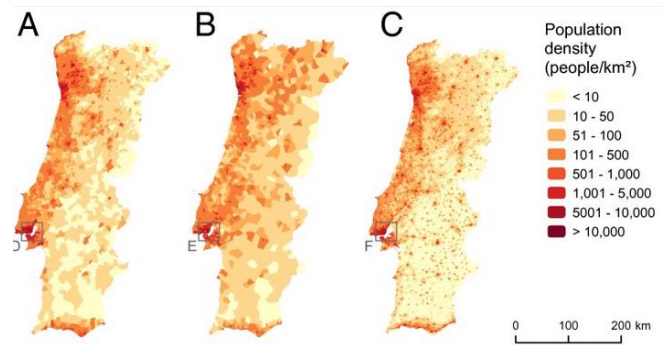


Рис.1. Плотность населения, рассчитанная по “телефонной” активности [17].

Изменение плотности во времени приводит к популярному способу визуализации – тепловым картам. И именно изменение плотности населения во времени – это отправная точка для транспортного моделирования. Это изменение плотности говорит о том, что какое-то количество абонентов переместилось из одного района в другой. Район в данном случае – это какая-то географически очерченная область, гео-квадрат. Размеры таких районов могут быть разными, минимальные размеры определяются техническими возможностями сети (характеристиками расположения базовых станций). Естественно, малые районы (данные по ним) могут быть агрегированы. В частности – и до каких-то административных единиц. В итоге – у нас есть пары районов (откуда-куда) и количество перемещающихся абонентов для каждой из пар. Такие перемещения мобильных абонентов, очевидно, должны быть как-то отображены в транспортной активности. Попросту – абоненты как-то добирались из района в район. Если районы перемещения не находятся в непосредственной близости (так что мы можем предположить пешеходный переход), то при перемещении, очевидно, использовался какой-то транспорт. Соответственно, система транспортного моделирования должна каким-то образом распределить (“рассадить”) этих пользователей по транспортным средствам. Здесь возможны различные варианты, в зависимости от того есть или нет дополнительная информация. Если, например, доступна информация по валидациям проездных документов в общественном транспорте, то можно выделить таких мобильных абонентов из общего числа перемещений, а оставшихся считать перемещающимися автомобильным транспортом. Если таковой информации нет, то можно исходить из максимально возможной загрузки общественного транспорта на данном направлении. При определении возможных режимов использования общественного транспорта исходят, как правило, из того, что пользователь выберет оптимальный по времени маршрут.

В целом, алгоритм расчета транспортных потоков можно представить следующим образом. Допустим, что у нас есть программный интерфейс для вычисления маршрута в городе между двумя точками. Типа того, что делает, например, Google Transit. Например, на рисунке

2 изображен пользовательский интерфейс системы. Здесь визуально показан маршрут между двумя точками в городе и предлагаемый к использованию транспорт. На рисунке 3 представлены детали данного маршрута. Весь функционал доступен и через программный интерфейс (API).

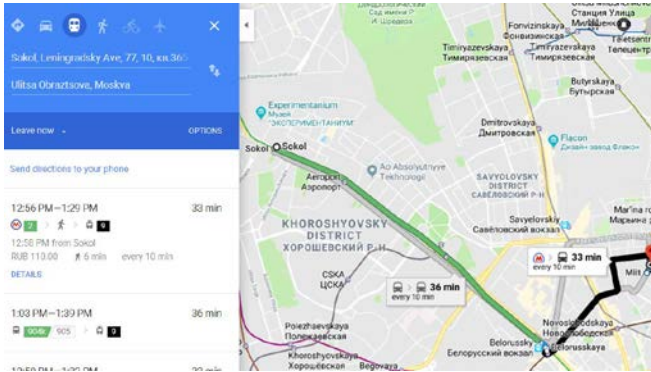


Рис. 2. Маршрут в Google Transit (пользовательский интерфейс)

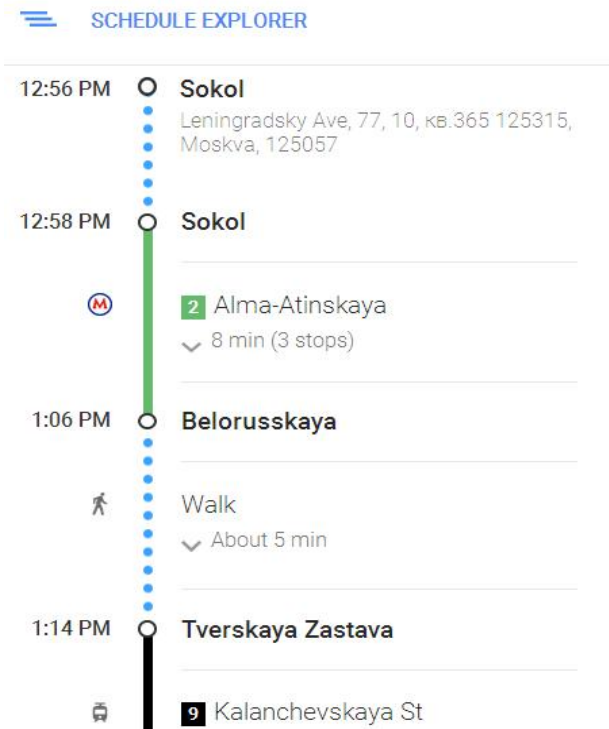


Рис. 3. Детали маршрута в Google Transit

Тогда алгоритм расчета транспортной загрузки будет выглядеть следующим образом:

Для каждой пары районов (гео-квадратов) $A > B$, где определено перемещение в момент времени t

Пусть N – количество перемещавшихся абонентов

Для каждого абонента из N

Определяем оптимальный маршрут $a > b$, где a есть середина квадрата A , а b – середина квадрата B

Для каждого вида транспорта, входящего в расшифровку маршрута

Обновить информацию о загрузке, с учетом времени, необходимого для предыдущих перемещений

Иными словами: расшифровка маршрута дает список видов транспорта, а также время, когда они будут использованы (относительно времени начала поездки). Это и позволяет спрогнозировать нагрузку на конкретные виды транспорта.

В работе [19] авторы используют модель на базе LDA для классификации перемещений пользователей, подобно тому, как классифицируются текстовые документы по темам. Для представления перемещений пользователя каждое индивидуальное местоположение изначально классифицируется по типу дом-работа. Далее с шагом 30 мин. записываются перемещения абонента в терминах введенной классификации (рис. 4).

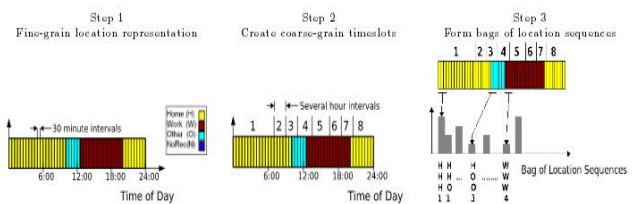


Рис.4. Классифицированные перемещения [19].

В таком случае все перемещения представляются как последовательности тегов. Вот эти последовательности и есть предмет анализа (кластеризации).

В работе [20] есть определение схожести перемещений (траекторий) мобильных абонентов. Метрики, которые вводятся для сравнения:

- Расстояние (Distance) между абонентами – определяется как физическое расстояние между двумя их наиболее часто встречающимися местоположениями
- Пространственная оценка местоположения (Spatial Co-Location Rate) двух абонентов – вероятность оказаться в одном месте, не обязательно в одно время
- Взвешенное пространственное косинусное сходство (Weighted Spatial Cosine Similarity) для частоты визитов
- Оценка местоположения (Co-Location Rate) – вероятность оказаться в одном месте в течение одного и того же временного интервала (часа)
- Взвешенная оценка местоположения (Weighted Co-Location Rate) - вероятность совпадения местоположения двух пользователей в пространстве в одно время, нормированная по плотности популяции в то же время

- Дополнительная оценка местоположения (Extra-role Co-Location Rate) - вероятность совпадения местоположения двух абонентов в течение часа в выходные дни.

Оригинальное (исходное) использование данной информации – восстановление скрытых связей между абонентами. Применительно к транспортным задачам, связь между абонентами может вести к схожему транспортному поведению - маршруты таких пользователей будут одинаковы. Это важно, поскольку как было сказано выше, основа для моделирования транспортных потоков – это матрица корреспонденции (в англоязычной литературе – Origin Destination или OD матрица).

Здесь хотелось бы отметить следующее. В просмотренных нами работах, матрица перемещений рассматривалась всегда в единственном варианте. Но очевидно, что количество перемещающихся из района в район зависит от времени дня. Перемещение начинается в какое-то определенное время. Например, утром из пригородов большого города едут на работу, вечером из города едут в пригород домой. Также очевидно, что для целей транспортного планирования необходимо понимать как изменение этой матрицы по времени в течение дня, так и изменение этих параметров по дням в течение некоторого календарного периода. Поскольку в некоторых случаях может быть сезонная составляющая в перемещениях, то календарный период в таком случае должен быть достаточно большим. Например, сравнение перемещений в будние и выходные дни может помочь в определении перемещений связанных с работой. Их можно было бы оценивать как более стабильные (постоянные) и, с большей вероятностью, привязанные ко времени. Представляется, что вопрос о статистических характеристиках OD матрицы и интерпретации полученных результатов является открытой темой.

Другим открытым направлением исследования является сравнение данных по перемещениям от операторов с известной информацией по использованию транспорта. Например, для пригородной станции железной дороги сравнить перемещения в районы, располагающиеся по направлению движения на железной дороге по данным операторов и информацию об использовании данной железнодорожной станции. Это позволило бы понять, например, какую долю перемещающихся абонентов забирает на себя конкретный вид транспорта (в данном случае – железная дорога). При этом было бы возможно построение какой-то визуальной характеристики (изохронной линии), которая показывала бы охватываемый перевозчиком процент абонентов уже в географической плоскости. Очевидно, что такого рода данные разнились бы от станции к станции, но, в то же время, это была бы качественная характеристика использования транспорта в данном районе. Можно заметить, что, например, исследования, касающиеся использования

телекоммуникационных данных для восстановления (поиска) связей между абонентами или шаблонов использования телефонов используют (обрабатывают) данные об индивидуальных операциях (именно они агрегируются или кластеризуются, например). В то время как транспортные модели начинаются уже с агрегированных, в какой-то степени, данных и теряют на этом моменте важную информацию.

Точно такой же вопрос об использовании (анализе) индивидуальных данных стоит и в основе построения самой матрицы корреспонденции. Какие мобильные абоненты должны выбираться для анализа перемещений? Все, кто на данный момент времени определен (классифицирован) как находящийся в данном районе? Или те, для кого данный район является домашним (рабочим)?

Другой момент, который теряется при агрегировании информации о перемещениях – это частотная информация. Итоговая цифра количества переместившихся за месяц, например, не дает никакого ответа на вопрос – а как были распределены эти перемещения среди абонентов по частоте?

В работе [21] рассматривается классическая модель восстановления матрицы OD, и использование этих данных при проектировании железной дороги. Отметим еще раз, что именно данные телекоммуникационных операторов серьезно изменили постановку задач для транспортных систем. То, что традиционно прогнозировалось (потоки), становится исходными данными. Восстановление матрицы корреспонденций (она же OD матрица) становится не нужным, поскольку ее также могут собирать телекоммуникационные операторы.

Какие же другие задачи решаются в этом случае? Один из интересных результатов (проект SOMOBIL) представлен в работе [22]. Это именно Фреймворк для транспортного планирования, основанный на данных телекоммуникационных операторов. Один из интересных результатов, который можно найти в данной работе – это восстановление (определение) того, каким транспортом пользовались мобильные абоненты при перемещении.

Классическая работа [23] посвящена выяснению шаблонов перемещения (мобильности). Авторы утверждают о наличии всего лишь 17 уникальных шаблонов, которых достаточно для описания перемещений 90% людей в мире. При этом шаблоны перемещений конкретного человека, чаще всего, остаются стабильными в течение месяцев. Авторы заключают, что ежедневная мобильность людей может быть воспроизведена моделированием марковских цепей для периодов высокочастотных поездок, за которыми следуют периоды более низкой активности в качестве основной модели.

Авторы отмечают, что современное общество и окружающая среда формируются под влиянием моделей мобильности людей в различных масштабах.

Долгосрочные и междугородние поездки, как правило, состоят из редких и редких событий, таких, как международные рейсы или поездки между городами. Напротив, краткосрочные поездки, в основном, состоят из внутригородских поездок, таких как поездка на работу или поездка за покупками. Эти поездки проходят с высокой регулярностью, как правило, в соответствии с суточным циркадным ритмом. Исследования мобильности людей в широком масштабе, мотивированные пониманием глобального распространения эпидемий, выявили интересные свойства, лежащие в основе моделей мобильности.

В настоящее время крупномасштабные модели мобильности людей описываются тремя общепринятыми показателями: распределение расстояния поездок, радиус гирации (как распределены места посещения) и количество посещенных мест во времени.

Подробный обзор моделей обработки CDR данных содержится в работе [24]. Авторы сосредоточились на описании моделей выявления социальных связей по данным телефонных взаимодействий. Это также имеет отношение к транспорту - мобильность между двумя пунктами зависит от расстояния между этими двумя пунктами и интенсивности коммуникаций между ними.

И, как последнюю работу в данном обзоре, отметим хорошо подобранную коллекцию работ по анализу траекторий [25]. Помимо собственно классификации исследований здесь есть интересная подборка доступных датасетов.

БИБЛИОГРАФИЯ

- [1] Horak R. Telecommunications and data communications handbook. – John Wiley & Sons, 2007.
- [2] Namiot D. et al. On the assessment of socio-economic effects of the city railway //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 1. – С. 92-103.
- [3] Namiot D., Pokusaev O., Lazutkina V. On passenger flow data models for urban railways //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 3. – С. 9-14.
- [4] Call detail record https://en.wikipedia.org/wiki/Call_detail_record Retrieved: Aug, 2018
- [5] Namiot D. Geo messages //Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2010 International Congress on. – IEEE, 2010. – С. 14-19.
- [6] Namiot D., Sneps-Snepe M. Where are they now—safe location sharing //Internet of Things, Smart Spaces, and Next Generation Networking. – Springer, Berlin, Heidelberg, 2012. – С. 63-74.
- [7] Social Dynamics: Signals and Behavior <https://www.media.mit.edu/publications/social-dynamics-signals-and-behavior/> Retrieved: Sep, 2018
- [8] Eagle, Nathan, and Alex Pentland. "Reality mining: sensing complex social systems." *Personal and ubiquitous computing* 10.4 (2006): 255-268
- [9] Namiot, Dmitry, and Manfred Sneps-Snepe. "On open source mobile sensing." *International Conference on Next Generation Wired/Wireless Networking*. Springer, Cham, 2014.
- [10] Незнанов, И. В., and Д. Е. Намиот. "Контроль транспортных маршрутов с помощью мобильных телефонов." *International Journal of Open Information Technologies* 3.8 (2015).
- [11] Rojas, Francisca, et al. "Real Time Rome." *Networks and Communication Studies* 20.3 (2008): 247-258.
- [12] Calabrese, Francesco, et al. "Real-time urban monitoring using cell phones: A case study in Rome." *IEEE Transactions on Intelligent Transportation Systems* 12.1 (2011): 141-151.
- [13] Намиот, Д. Е., & Колосова, А. И. (2013). Об определении владельцев мобильного телефона. *International Journal of Open Information Technologies*, 1(8).
- [14] Namiot, Dmitry, and Manfred Sneps-Snepe. "On software standards for smart cities: API or DPL." *ITU Kaleidoscope Academic Conference: Living in a converged world-Impossible without standards?*. Proceedings of the 2014. IEEE, 2014.
- [15] N. Eagle, A. S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009
- [16] Candia, Julián, et al. "Uncovering individual and collective human dynamics from mobile phone records." *Journal of physics A: mathematical and theoretical* 41.22 (2008): 224015.
- [17] Deville, Pierre, et al. "Dynamic population mapping using mobile phone data." *Proceedings of the National Academy of Sciences* 111.45 (2014): 15888-15893.
- [18] Aurenhammer, Franz, and Herbert Edelsbrunner. "An optimal algorithm for constructing the weighted Voronoi diagram in the plane." *Pattern Recognition* 17.2 (1984): 251-257.
- [19] Katayoun Farrahi and Daniel Gatica-Perez. "Discovering routines from large-scale human locations using probabilistic topic models". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 2.1 (2011), p. 3.
- [20] Wang D. et al. Human mobility, social ties, and link prediction //Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. – Acm, 2011. – С. 1100-1108.
- [21] ANALYZING CELL PHONE LOCATION DATA FOR URBAN TRAVEL: CURRENT 3 METHODS, LIMITATIONS AND OPPORTUNITIES http://blogs.worldbank.org/sites/default/files/2014_trb_cdr_boston_ri_o.pdf
- [22] Elias, Daniel, et al. "SOMOBIL—improving public transport planning through mobile phone data analysis." *Transportation Research Procedia* 14 (2016): 4478-4485.
- [23] Schneider, Christian M., et al. "Unravelling daily human mobility motifs." *Journal of The Royal Society Interface* 10.84 (2013): 20130246.
- [24] Blondel, Vincent D., Adeline Decuyper, and Gautier Krings. "A survey of results on mobile phone datasets analysis." *EPJ data science* 4.1 (2015): 10.
- [25] Lohan, Elena-Simona, Tomi Kauppinen, and Sree Bash Chandra Debnath. "A survey of people movement analytics studies in the context of smart cities." 2016 19th Conference of Open Innovations Association (FRUCT). IEEE, 2016.

Use of telecommunications operators' data in transport planning

Dmitry Namiot, Oleg Pokusaev, Alexander Chekmarev

Abstract— The article deals with issues related to the use of data from telecommunications operators in transport planning. Recently, the penetration of mobile phones has ensured that it is the data collected by telecommunications operators becoming the main tool for measuring the movement of people in cities. It is no exaggeration to say that digital urban planning started with such data. It is that telecommunications operators naturally (for their own purposes of billing communication services) collect information about the presence of mobile devices in different areas of service. More precisely - about the service of mobile devices by different base stations of the operator, each of which is tied to a certain geographical area. Accordingly, at the operator level, it is clear when a particular mobile device has moved from one area to another (moved to service from one base station to another). These anonymous and time aggregated data provide information on the number of mobile devices (the number of owners of these devices) that have moved from one area to another in a given time interval. For example, in 15 minutes, 30 minutes, one hour, etc. This representation of human flows is the basis for transport planning.

Keywords— digital urban planning, transport planning, modeling, forecasting.

REFERENCES

- [1] Horak R. Telecommunications and data communications handbook. – John Wiley & Sons, 2007.
- [2] Namiot D. et al. On the assessment of socio-economic effects of the city railway //International Journal of Open Information Technologies. – 2018. – T. 6. – #. 1. – S. 92-103.
- [3] Namiot D., Pokusaev O., Lazutkina V. On passenger flow data models for urban railways //International Journal of Open Information Technologies. – 2018. – T. 6. – #. 3. – S. 9-14.
- [4] Call detail record https://en.wikipedia.org/wiki/Call_detail_record Retrieved: Aug, 2018
- [5] Namiot D. Geo messages //Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2010 International Congress on. – IEEE, 2010. – S. 14-19.
- [6] Namiot D., Sneps-Snepp M. Where are they now—safe location sharing //Internet of Things, Smart Spaces, and Next Generation Networking. – Springer, Berlin, Heidelberg, 2012. – S. 63-74.
- [7] Social Dynamics: Signals and Behavior <https://www.media.mit.edu/publications/social-dynamics-signals-and-behavior/> Retrieved: Sep, 2018
- [8] Eagle, Nathan, and Alex Pentland. "Reality mining: sensing complex social systems." Personal and ubiquitous computing 10.4 (2006): 255-268
- [9] Namiot, Dmitry, and Manfred Sneps-Snepp. "On open source mobile sensing." International Conference on Next Generation Wired/Wireless Networking. Springer, Cham, 2014.
- [10] Neznanov, I. V., and D. E. Namiot. "Kontrol' transportnyh marshrutov s pomoshh'ju mobil'nyh telefonov." International Journal of Open Information Technologies 3.8 (2015).
- [11] Rojas, Francisca, et al. "Real Time Rome." Networks and Communication Studies 20.3 (2008): 247-258.
- [12] Calabrese, Francesco, et al. "Real-time urban monitoring using cell phones: A case study in Rome." IEEE Transactions on Intelligent Transportation Systems 12.1 (2011): 141-151.
- [13] Namiot, D. E., & Kolosova, A. I. (2013). Ob opredelenii vladel'cev mobil'nogo telefona. International Journal of Open Information Technologies, 1(8).
- [14] Namiot, Dmitry, and Manfred Sneps-Snepp. "On software standards for smart cities: API or DPI." ITU Kaleidoscope Academic Conference: Living in a converged world-Impossible without standards?, Proceedings of the 2014. IEEE, 2014.
- [15] N. Eagle, A. S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences , 106(36):15274–15278, 2009
- [16] Candia, Julián, et al. "Uncovering individual and collective human dynamics from mobile phone records." Journal of physics A: mathematical and theoretical 41.22 (2008): 224015.
- [17] Deville, Pierre, et al. "Dynamic population mapping using mobile phone data." Proceedings of the National Academy of Sciences 111.45 (2014): 15888-15893.
- [18] Aurenhammer, Franz, and Herbert Edelsbrunner. "An optimal algorithm for constructing the weighted Voronoi diagram in the plane." Pattern Recognition 17.2 (1984): 251-257.
- [19] Katayoun Farrahi and Daniel Gatica-Perez. "Discovering routines from large-scale human locations using probabilistic topic models". In: ACM Transactions on Intelligent Systems and Technology (TIST) 2.1 (2011), p. 3.
- [20] Wang D. et al. Human mobility, social ties, and link prediction //Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. – Acm, 2011. – S. 1100-1108.
- [21] ANALYZING CELL PHONE LOCATION DATA FOR URBAN TRAVEL: CURRENT 3 METHODS, LIMITATIONS AND OPPORTUNITIES http://blogs.worldbank.org/sites/default/files/2014_trb_cdr_boston_rio.pdf
- [22] Elias, Daniel, et al. "SOMOBIL—improving public transport planning through mobile phone data analysis." Transportation Research Procedia 14 (2016): 4478-4485.
- [23] Schneider, Christian M., et al. "Unravelling daily human mobility motifs." Journal of The Royal Society Interface 10.84 (2013): 20130246.
- [24] Blondel, Vincent D., Adeline Decuyper, and Gautier Krings. "A survey of results on mobile phone datasets analysis." EPJ data science 4.1 (2015): 10.
- [25] Lohan, Elena-Simona, Tomi Kauppinen, and Sree Bash Chandra Debnath. "A survey of people movement analytics studies in the context of smart cities." 2016 19th Conference of Open Innovations Association (FRUCT). IEEE, 2016.