

Куда и когда – об одном подходе к анализу трафика в городе

Д.Е. Намиот, А.В. Акимов, М.Н. Некраплённая, О.Н. Покусаев

Аннотация— В этой статье речь идет об одной модели анализа мобильности в городах. Традиционно, при анализе перемещений используется временной домен. Это связано как с традиционными моделями анализа расписаний, так и с классическим подходом к представлению транспортных задач как задач прогнозирования трафика. Вместе с тем, развитие телекоммуникационных технологий и проникновение смартфонов привели к тому, что по перемещениям мобильных устройств можно достаточно точно измерять транспортные потоки. Соответственно, предсказание транспортных потоков становится не самой актуальной задачей – нет необходимости предсказывать то, что точно измеряется. В современных условиях, данные о потоках становятся метрикой, которая отражает процессы (ситуации) в городе. Например, данные о перемещениях между станциями метро показывают шаблоны использования соответствующих станций, которые, на самом деле, описывают модели функционирования прилегающих территорий: является ли это спальным районом, где люди уезжают утром на работу и возвращаются вечером, как различаются режимы работы по будням и выходным дням и т.д. И любые изменения в таких шаблонах будут сигналом об изменении режимов использования. Или, другими словами, являться отражением каких-то процессов в городе. В данной статье речь идет об одном из подходов к анализу трафика, связанному как раз с поиском шаблонов перемещения.

Ключевые слова—мобильность, Умный Город.

I. ВВЕДЕНИЕ

Одной из наиболее часто используемых форм описания мобильности (представления перемещений) являются так называемые матрицы корреспонденции. В англоязычной литературе используется аббревиатура OD – матрицы (origin – destination). В таких матрицах показывается количество перемещений между двумя пунктами, представленными в строках и столбцах матрицы. Такие перемещения относятся, естественно, к некоторому временному интервалу, который может быть, вообще говоря, произвольным: 15 мин, 30 мин, 1 час, 1 день и т.д. Соответственно, в матрице корреспонденций M , элемент m_{ij} показывает количество

перемещений за заданный временной интервал t от пункта i до пункта j .

В качестве пунктов могут выступать, например, железнодорожные станции (станции городской железной дороги), станции метро или просто некоторые описанные географические области. Например, для Москвы (московского региона) такие матрицы существуют для квадратов географической сетки (георешеток) со стороной 500 м [1].

Такие матрицы корреспонденции играют фундаментальную роль в представлении (описании) трафика, поскольку дают исчерпывающую информацию о перемещениях. Во множестве транспортных задач именно восстановление (построение, прогнозирование) матрицы перемещений является основной задачей [2].

В каких-то условиях функционирования транспортных систем информация о матрицах корреспонденции может собираться автоматически. Если пассажиры должны отмечать начало и конец поездки (регистрировать свои проездные документы в начале и конце поездки), то это и будет соответствовать сбору матрицы корреспонденции. Так, например, организованы перевозки на пригородных электричках в московском регионе. В случае отсутствия отметок о завершении поездки (например, городской транспорт в Москве), восстановлению матриц корреспонденции может помочь использование стандартных платежных инструментов для оплаты проезда. Например, для Москвы таким платежным инструментом для транспорта является карта Тройка. Модель рассуждений здесь может быть следующая. Если зафиксировано использование некоторой карты на станции $C1$ (проход), а далее, после некоторого временного перерыва, на станции $C2$ (второй проход), то мы предполагаем наличие поездки ($C1, C2$) (пассажир, фактически, проехал по маршруту $C1 - C2$, пробыл в точке назначения некоторое время и начал новый маршрут в $C2$) [3].

Но основным инструментом сбора информации о перемещениях, как отмечалось ранее, в современных условиях являются мобильные телекоммуникационные операторы [4]. Именно они, фактически, способны собирать информацию обо всех перемещениях своих абонентов, которые и являются пассажирами общественного транспорта или водителями личных автомобилей. Основой для такой информации является переключение мобильного устройства абонента между различными базовыми станциями, местоположение (координаты) которых известны. При этом данные

Статья получена 15 января 2021.

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

А.В. Акимов - Департамент транспорта и развития дорожно-транспортной инфраструктуры города Москвы (email: akimov_post@mail.ru).

М.Н. Некраплённая – МГУ имени М.В. Ломоносова (email: maria240398@mail.ru).

О.Н. Покусаев - РУТ (МИИТ) (email: o.pokusaev@rut.digital).

могут собираться как относительно квадратов географической сетки, так и, например, относительно станций метрополитена. В последнем случае используется факт поддержки связи в метро. Мобильный телефон пассажира при спуске в метро переключается на базовую станцию в метро. Так определяется станция метро, где пассажир вошел. Соответственно, по переключению на другую базовую станцию, можно определить, где пассажир вышел [5].

Отметим также, что сбор и анализ матриц корреспонденции не имеет ничего общего с отслеживанием пассажиров. Такие матрицы полностью анонимны. Более того, при их сборе операторы всегда суммируют перемещения за некоторый интервал времени, так что одиночные перемещения никогда не представляются (spatial cloaking) [6].

В настоящей статье речь будет идти о поездках метро, но дело здесь только в простоте и размере исследуемых данных. Все соображения, рассматриваемые в данной статье, остаются такими же и для остальных матриц корреспонденции.

II. АНАЛИЗ МАТРИЦ КОРРЕСПОНДЕНЦИИ

Как может использоваться информация о потоках пассажиров по станциям? На самом деле, работ, анализирующих именно матрицы корреспонденции не так и много. Это связано как раз с высказанными выше соображениями о том, что в транспортных системах рассматривается как раз построение (прогнозирование) матриц корреспонденции, а не анализ таких данных.

Во-первых, самое простое – это расчет возможных перераспределений пассажирских потоков. Такие данные могут использоваться, например, при вводе новых станций (маршрутов) в дополнение к существующим или наоборот, закрытии каких-либо станций (маршрутов). Исходя из того, что пассажиры будут выбирать наиболее короткие по времени маршруты можно рассчитать, кто (сколько) из существующих пассажиров изменит свои маршруты. Но это вряд ли можно назвать прогнозом, поскольку расчеты здесь абсолютно детерминированные.

Эти расчеты, конечно, отличаются от полностью статически расчетов типа анализа центральности станций [2], но, в целом, не представляют большой проблемы. Такая информация будет полезна, например, при проектировании новых пересадочных узлов, объединении метро и городской железной дороги. Однако при таком подходе теряется (не учитывается) возможное распределение пассажиров по времени. Очевидно, что в некоторых случаях недостаточно просто указать новое количество пассажиров, которое появится на станции. Смысл в том, что эти новые пассажиры появятся не одномоментно, а будут как-то распределены по времени. И именно эта информация о распределении также может быть существенной. Например, новые пассажиры, добавленные к существующему утреннему пику, могут оказаться критичным изменением, превышающим пропускную

способность.

Следующее замечание относится к тому, что необходимо отметить – это замечание о том, что в описанном выше представлении матрица корреспонденций представляет собой замкнутую систему. Это означает, что в матрице, например, никоим образом не отражается информация о добавлении новых пассажиров в систему (например, новый маршрут из пригорода до станции метро). В матрице нет информации о каких-то разовых событиях (концерты, футбольные матчи и т.п.), которые разово влияют на поведение пассажиров. Матрица корреспонденции отражает результаты таких изменений, но причина изменений, естественно, отсутствует.

В работе [2] мы приводили пример с переключением пригородных автобусов с автовокзала около метро Тушинская на новый автовокзал около метро Ховрино, который вызвал изменения в утренней загрузке на зеленой линии метро. Изменение можно было явно зафиксировать, но предсказать, естественно, было невозможно. Это была только фиксация постфактум, поскольку вся возможная объясняющая информация является внешней по отношению к матрице корреспонденции.

Отсюда и следовало наше предложение, считать матрицу измерительным инструментом (метрикой), которая отражает текущее поведение пассажиров. Матрица корреспонденций, полученная описанным выше способом – это некоторый измерительный инструмент, с помощью которого можно определять наступление каких-то событий в городе или проверять результаты каких-либо действий [2]. Изменения в поведении пассажиров (изменение маршрутов, времени поездок) – это то, что можно определить с помощью анализа матриц корреспонденции. Каждое такое изменение должно, естественно, быть как-то объяснено с точки зрения каких-то изменений в городе. В работе [7] это называлось урбанистическими событиями. Например, эффект от проведения каких-либо массовых событий будет отображаться в транспортных перемещениях как избыточные (по сравнению с некоторым предшествующим периодом) входы и выходы на определенных станциях (участникам нужно приехать и уехать на мероприятия).

Соответственно, анализ матриц корреспонденции должен обслуживать, в первую очередь, тех, кто занимается эксплуатацией транспорта, организацией движения и городским планированием. Именно для них, основной вопрос – это пространственно-временные характеристики потока. Как он распределен на конкретной станции по входу и выходу в течение дня, остается ли такое распределение стабильным, как определить выбросы (аномалии), как понять, что бывшие аномалии стали новой нормой и т.д. Это все имеет самое прямое отношение к тому, что в Умном городе называется мобильность как сервис (мобильность как услуга). Для представления сервиса, естественно, необходимо понимать, как этот сервис

(спрос на услугу) устроен [2].

При этом необходимо отметить, что на сегодняшний день нельзя предложить какую-то универсальную группу метрик для оценки матриц корреспонденции. Весь наш предыдущий анализ показывает, что такого рода универсальных индикаторов нет. Хотя работы по их поиску ведутся [8, 9, 10].

В предыдущих работах мы описывали собственные предложения, базирующиеся на анализе литературы и предшествующих разработках, относительно того, что могло бы быть использовано для описания характеристик таких матриц.

Во-первых, это, конечно, распределение входов/выходов по времени. Причины важности понимания этих параметров очевидны. Пики в этих распределениях должны укладываться в пропускные способности станций. Работа наземного транспорта должна быть как-то синхронизована с потоками пассажиров. Последовательность нескольких станций с одинаковым утренним пиком по входу, например, может привести к невозможности (затруднению) входа на каких-то станциях. Это означает, что такие распределения должны рассматриваться совместно. Тесно связанный с этим вопрос – это аномалии во временных распределениях входов и выходов. Это как раз оценка эффектов от каких-либо событий в городе. Технически – это поиск аномалий во временных рядах.

Может представлять интерес исследование зависимости в трафике между станциями. Очевидно, что для большой системы, которой является московское метро, разовое увеличение количества пассажиров, вошедших на конечной станции на севере города, будет “не замеченным” где-то на юге. Но на ближайших станциях это будет, конечно, заметно. По сути – это некоторая структурная характеристика потоков – как связаны потоки на разных станциях. Эти зависимости разные, что проверилось и в наших работах.

В общем, изложенные выше метрики можно представить как структурные описания характеристик использования станций. Соответственно, в финальной точке, мы хотим создать систему, которая сможет описывать характер использования станций, определять аномалии, находить моменты времени, когда по каким-то причинам изменились режимы использования станций (перегонов) и т.д.

Результатом работы является, например, зонирование станций метрополитена по характеру их использования [5]. Эта картина отражает схожесть различных станций по моделям (характеру) их использования. При этом (это важно для дальнейшего рассмотрения) характер использования определяется именно во временном домене. Это весьма понятный для объяснения способ. Одинаковыми в смысле характеристик использования оказывались, например, станции, которые работали, в основном, на вход утром и выход вечером (характерно для спальных районов), либо наоборот – на выход утром

и вход вечером (офисные районы) и т.д. Это проиллюстрировано на рисунке 1.

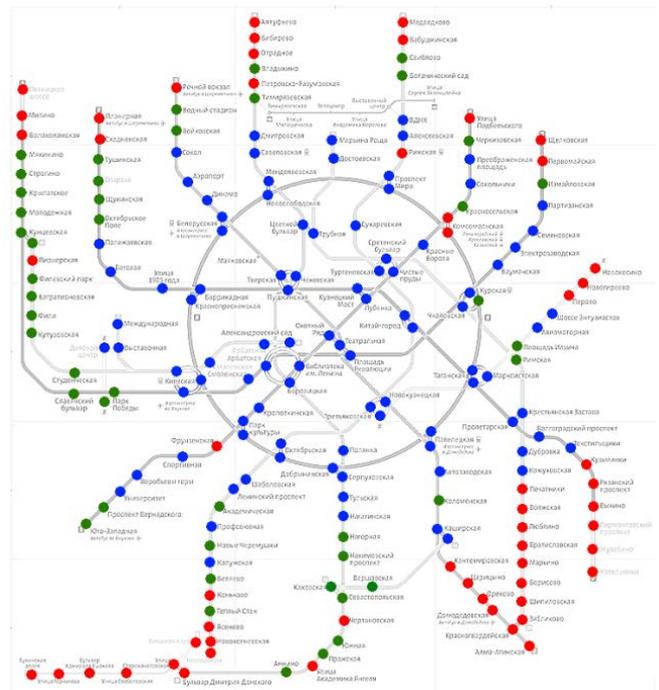


Рис. 1. Зонирование станций [5]

В настоящей же работе мы хотели поставить вопрос о применимости альтернативной модели. Что если исследовать перемещения в пространственном разрезе? То есть, не “когда ездят со станции/на станцию”, а “куда ездят со станции/откуда ездят на станцию”?

III. АНАЛИЗ ПЕРЕМЕЩЕНИЙ В ПРОСТРАНСТВЕННОМ РАЗРЕЗЕ

Сама идея исследования именно пространственных характеристик (поездки, маршруты) пришла из анализа перемещений с использованием пригородных железных дорог [11] и работ по прогнозированию загрузки МЦД [12]. Для пригородных железных дорог существовал четкий шаблон выбора конечной станции маршрута как первой станции, где есть пересечение с линией (линиями) метро. И этот шаблон менялся в выходные дни. Собственно говоря, новая схема движения на МЦД должны была изменить этот шаблон и предложить пассажирам продолжать использовать железную дорогу в городе, а не пересаживаться на метро при первой же возможности.

Также пространственный анализ приводит к оценке длительности перемещений (длительности маршрутов), а это, как известно, одна из основных характеристик транспортного поведения. Другое соображение состоит в том, что конечные пункты маршрутов должны соответствовать тому, что принято называть точками притяжения. При этом такие точки притяжения будут очевидно, зависеть от времени и, возможно, от дней недели. Например, станции, где завершается большинство поездок (маршрутов) утром – это то, что

во временном разрезе исследований соответствовало рабочим районам и т.д.

Что именно мы хотим исследовать таким образом? Представляется, что на первом шаге будет разумно исследовать вопрос о том, существуют ли какие-то шаблоны в таких перемещениях. То есть, мы хотим сохранить предыдущую идею о поиске шаблонов, которые нужны для того, чтобы по их изменению сигнализировать о каких-то событиях в городе.

Очевидно, что в данном случае речь не идет о группе станций. Такого рода исследования нужно будет проводить для конкретных станций. Возможно, что такого рода шаблоны будут существовать не для всех станций и не для всех временных интервалов.

Там же, где это, возможно, существует, под шаблоном будем понимать следующее. Для станции A и внутрисуточного временного интервала t мы можем получить список станций, куда отправлялись пассажиры:

$$P(A, t) \rightarrow \{S_i\}$$

такой список может быть упорядочен, например, по убыванию количества пассажиров, отправившихся на соответствующую станцию.

Наличие шаблона будет определяться тем, что такой же список $P(A, t)$ для других дней (других рабочих дней, других выходных или праздничных дней – в рабочие и выходные дни режимы перемещений могут быть, очевидно, разными) будет совпадать с исходным.

Метрики для оценки совпадений можно выбирать (пробовать) разные. Например, вычислять расстояние Левенштейна для последовательностей станций, отобрать несколько первых станций назначения по пассажиропотоку и сравнивать такие множества с использованием коэффициента Жаккара. Можно вместо станций назначения использовать время проезда до них, и сравнивать тогда маршруты по средне-взвешенному (с учетом пассажиропотока) времени.

Какой “урбанистический” смысл может быть в таком выделении шаблонов? Когда мы говорили о временном домене, кластеризация станций отражала в той или иной степени рабочее расписание для людей, связанных по месту жительства и/или работы с соответствующими станциями. Кластеризация же маршрутов (если таковая будет обнаружена) соответствует же точкам притяжения трафика, то есть оценке того, какие станции собирают больше пассажиров. При этом такая информация нам интересна не в форме общей статистике, а в привязке ко времени дня. Как и с определениями кластеров во временном аспекте, изменения в предпочтениях для путей (маршрутов) будут сигнализировать об изменениях в городе (изменениях в точках притяжения

трафика и/или режимов их использования).

В качестве первых станций-кандидатов на поиск перемещений было бы интересно рассмотреть станции, которые являются “точками входа” в систему метро – там, где находятся автобусные станции для пригородных направлений, а также станции, где есть пересадка с МЦД. Для таких станций интересно было бы также сравнить маршруты в рабочие дни и выходные.

Для простой проверки гипотезы о существовании стабильных маршрутов были выполнены простые расчеты по популярности конечных станций в поездах со станции Речной вокзал в феврале 2018 года. Ниже приведены данные о том, сколько раз указанная конечная станция оказалась самой популярной точкой назначения:

Войковская	22
Тверская	4
Китай-город	1
Маяковская	1

Как видно, все практически предопределено, большинство маршрутов: Речной вокзал – Войковская. Интересно, также, что центральные станции (Тверская, Китай-город) появляются в выходные дни. Такая же картина видна и на получасовых интервалах. В рабочие дни в интервале 6:00 – 8:30 также лидирует маршрут Речной вокзал – Войковская. А для самой станции Войковская, начиная с 13:30, основным маршрутом является Войковская – Речной вокзал.

IV. ЗАКЛЮЧЕНИЕ

Настоящая статья является продолжением серии исследований, посвященной анализу матриц корреспонденции. Общим для них является то, что от задач прогнозирования необходимо переходить к поведенческому анализу и выявлению шаблонов использования транспортной сети. Таковые шаблоны всегда отражают текущую ситуацию в городе, тогда как отклонения от них – это результат реакции на какие-то разовые события, либо переключение на новую “нормальность”, то есть, фактически – смена шаблонов. Матрица корреспонденции в этом случае будет выступать как индикатор и средство измерения при обнаружении и оценке событий с точки зрения вовлеченности людей. В предыдущих работах поведение пассажиров на станциях метро (фактически - режимы использования станции) определялось распределениями входов и выходов по времени в течение дня, распределением соотношений входов и выходов в течение дня, а также внутренними характеристиками распределений входов и выходов (максимумами и минимумами). Это соответствовало анализу во временном домене. С технической точки зрения – кластеризации временных рядов. Данная статья является первой работой, в которой рассматривается постановка задачи анализа трафика в пространственном домене. Куда именно ездят в городе с определенных станций,

если ли здесь какие-то закономерности, как это зависит от времени и дня недели и т.д. – вот типичные вопросы, ответы на который должен давать такой анализ.

Нам представляется, что такая постановка задач анализа трафика представляет значительный интерес. В первую очередь, за счет того, что анализ направлений приводит к оценке длительности поездок, а это есть одна из важнейших характеристик транспортного поведения.

БИБЛИОГРАФИЯ

- [1] Misharin, A., D. Namiot, and O. Pokusaev. "On Processing of Correspondence Matrices in Transport Systems." 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). IEEE, 2019.
- [2] Namiot, Dmitry, et al. "OD-matrix and passenger flow analysis." International Journal of Open Information Technologies 8.4 (2020): 25-30.
- [3] Namiot, Dmitry, and Manfred Sneps-Sneppe. "A Survey of Smart Cards Data Mining." AIST (Supplement). 2017.
- [4] Bulygin, Mark, and Dmitry Namiot. "On the possibilities of using the data of cellular operators to solve the problems of digital urbanism." International Journal of Open Information Technologies 9.1: 48-57.
- [5] Nekraplonna, Mariia, and Dmitry Namiot. "Metro correspondence matrix analysis." International Journal of Open Information Technologies 7.7 (2019): 68-80.
- [6] Niu, Ben, et al. "A novel attack to spatial cloaking schemes in location-based services." Future Generation Computer Systems 49 (2015): 125-132.
- [7] Bulygin, Mark, and Dmitry Namiot. "Anomaly Detection Method For Aggregated Cellular Operator Data." 2021 28th Conference of Open Innovations Association (FRUCT). IEEE, 2021.
- [8] Yang, Chao, Fen Fan Yan, and Xiang Dong Xu. "Clustering Daily Metro Origin-Destination Matrix in Shenzhen China." Applied Mechanics and Materials. Vol. 743. Trans Tech Publications Ltd, 2015.
- [9] Duan, Zhengyu, et al. "Understanding multiple days' metro travel demand at aggregate level." IET Intelligent Transport Systems 13.5 (2018): 756-763.
- [10] Yang, Chao, Fenfan Yan, and Satish V. Ukkusuri. "Unraveling traveler mobility patterns and predicting user behavior in the Shenzhen metro system." Transportmetrica A: Transport Science 14.7 (2018): 576-597.
- [11] Namiot, Dmitry, Oleg Pokusaev, and Vasily Kupriyanovsky. "On railway stations statistics in Smart Cities." International Journal of Open Information Technologies 7.4 (2019): 19-24.
- [12] Pokusaev, Oleg, Alexander Chekmarev, and Dmitry Namiot. "City railways—who are their passengers?." IOP Conference Series: Earth and Environmental Science. Vol. 534. No. 1. IOP Publishing, 2020.

Where and when - about one approach to traffic analysis in the city

Dmitry Namiot, Andrey Akimov, Mariia Nekraplonna, Oleg Pokusaev

Abstract— This article deals with one model for analyzing urban mobility. Traditionally, the time domain is used in the analysis of movements. This is due to both traditional models of scheduling analysis and the classical approach to representing transport problems as traffic forecasting problems. At the same time, the development of telecommunication technologies and the penetration of smartphones have led to the fact that the movements of mobile devices can accurately measure traffic flows. Accordingly, the prediction of traffic flows is not the most urgent task - there is no need to predict what is being measured accurately. In modern conditions, data on flows are becoming a metric that reflects the processes (situations) in the city. For example, the data on movement between metro stations shows patterns of the use of the corresponding stations, which, in fact, describe the models of the functioning of the adjacent territories: is it a sleeping area where people leave in the morning to work and return in the evening, how are the working hours different on weekdays and weekends? days, etc. And any changes to such templates will signal a change in usage modes. Or, in other words, to be a reflection of some processes in the city. In this article, we are talking about one of the approaches to traffic analysis related to the search for movement patterns.

Keywords— mobility, Smart City.

REFERENCES

- [1] Misharin, A., D. Namiot, and O. Pokusaev. "On Processing of Correspondence Matrices in Transport Systems." 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). IEEE, 2019.
- [2] Namiot, Dmitry, et al. "OD-matrix and passenger flow analysis." International Journal of Open Information Technologies 8.4 (2020): 25-30.
- [3] Namiot, Dmitry, and Manfred Sneps-Snepp. "A Survey of Smart Cards Data Mining." AIST (Supplement). 2017.
- [4] Bulygin, Mark, and Dmitry Namiot. "On the possibilities of using the data of cellular operators to solve the problems of digital urbanism." International Journal of Open Information Technologies 9.1: 48-57.
- [5] Nekraplonna, Mariia, and Dmitry Namiot. "Metro correspondence matrix analysis." International Journal of Open Information Technologies 7.7 (2019): 68-80.
- [6] Niu, Ben, et al. "A novel attack to spatial cloaking schemes in location-based services." Future Generation Computer Systems 49 (2015): 125-132.
- [7] Bulygin, Mark, and Dmitry Namiot. "Anomaly Detection Method For Aggregated Cellular Operator Data." 2021 28th Conference of Open Innovations Association (FRUCT). IEEE, 2021.
- [8] Yang, Chao, Fen Fan Yan, and Xiang Dong Xu. "Clustering Daily Metro Origin-Destination Matrix in Shenzhen China." Applied Mechanics and Materials. Vol. 743. Trans Tech Publications Ltd, 2015.
- [9] Duan, Zhengyu, et al. "Understanding multiple days' metro travel demand at aggregate level." IET Intelligent Transport Systems 13.5 (2018): 756-763.
- [10] Yang, Chao, Fenfan Yan, and Satish V. Ukkusuri. "Unraveling traveler mobility patterns and predicting user behavior in the Shenzhen metro system." Transportmetrica A: Transport Science 14.7 (2018): 576-597.
- [11] Namiot, Dmitry, Oleg Pokusaev, and Vasily Kupriyanovsky. "On railway stations statistics in Smart Cities." International Journal of Open Information Technologies 7.4 (2019): 19-24.
- [12] Pokusaev, Oleg, Alexander Chekmarev, and Dmitry Namiot. "City railways—who are their passengers?." IOP Conference Series: Earth and Environmental Science. Vol. 534. No. 1. IOP Publishing, 2020.