

Социальные кластеры в городе

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

Аннотация— Цифровые методы регистрации поездок на транспорте позволяют сегодня довольно точно измерять транспортное поведение. Цифровые инструменты оплаты и развитие телекоммуникаций позволяют, в частности, иметь полную информацию о поездках пассажиров в городе. Соответственно, данные о перемещениях пассажиров становятся индикатором процессов, которые происходят в городе. Фактически, транспортные данные выступают в роли сенсоров, измерения которых отражают процессы в городской системе. Например, распределение входов и выходов конкретной станции метро отражают режим использования прилегающей территории. В настоящей работе рассматривается задача оценки социальных кластеров в городе на основе информации об использовании станций метро. В качестве основы для такой кластеризации используется информация о форме оплаты поездок.

Ключевые слова—метро, оплата поездок, транспортные данные

I. ВВЕДЕНИЕ

Социальное деление городских районов представляет собой достаточно обычное дело [1]. Чаще всего это бывает связано со стоимостью жилья, уровнем благосостояния, происхождением жителей и т.д. В России, например, как многократно отмечалось в литературе, основой благополучия того или иного района остается инфраструктура - транспортная, социальная, торговая [2,3]. Практически всегда локации без хорошей транспортной доступности оказываются неблагоприятными. Именно транспортная доступность, в первую очередь, отмечается как качество локации. Отсутствие транспортной доступности по принципу домино тянет вниз за собой все остальные составляющие, влияющие на развитие и оценки любой территории [4]. Транспортная доступность в таком большом городе как Москва – это практически всегда метро (или городские железные дороги, которые в последнее время интегрированы с метрополитеном в единое целое). Открытие новой станции метро “оживляет” окружающие районы и способствует процессу освоения любой городской территории.

В данной работе мы рассматриваем, в некоторой степени, обратную задачу. Идея данной работы состоит в том, чтобы по характеристикам использования

метрополитена получить оценку социальной среды для окружающего района. Сама идея использования транспортных данных именно как метрики (набора метрик) в Умном городе является вполне рабочей. Смысл ее в том, что цифровизация практически всех процессов в городе позволяет собирать множество данных, относящихся к использованию транспортной системы, большая часть которых еще сравнительно недавно можно было получить только в качестве прогноза. Теперь же становится бессмысленно прогнозировать то, что может быть точно измерено. Соответственно, “транспортные” данные приобретают другое назначение [5, 6]. Характеристики использования транспортной системы становятся метриками, которые отражают процессы в городе. Например, мы точно можем измерить трафик на входе и выходе каждой станции. Утренние пики на входе соответствуют “спальному” району, откуда утром едут на работу. Утренние пики по выходу соответствуют рабочему (офисному) району. То есть характеристики транспортного потока описывают разделение (или использование) города. Как это может быть использовано городскими службами? Изменение такого разделения (например, изменение режима использования станции, увеличение или уменьшение потока, изменение распределения по времени и т.д.) отражает какие-то процессы (постоянные или временные) в городе. Например, заселение нового жилищного комплекса, начало работы нового бизнес-центра или наоборот – закрытие старого. Временные пики в трафике могут быть привязаны к каким-то разовым событиям – футбольный матч, концерт и т.п. Во всех случаях именно изменение сложившегося режима использования (шаблона использования станции метро) отражает какую-то последовательность событий. Естественно, что изменение не определяет однозначно причину – это уже дело городских служб, но с метриками им легче будет искать основания. Транспортные метрики таким образом становятся некоторой универсальной системой измерения в городе.

II. ИСТОЧНИКИ ДАННЫХ

Что является источником транспортных данных? В данном случае, мы не рассматриваем системы краудсенсинга, когда какие-либо данные собираются непосредственно на клиентских устройствах. Точно также не используются данные социальных сетей и любых иных подобных источников.

В работе используются два основных источника

Статья получена 22 декабря 2021.

О.Н. Покусаев – РУТ (МИИТ); РОАТ (email: o.pokusaev@rut.digital)

Д.Е. Намиот – МГУ имени М.В. Ломоносова (email: dnamiot@gmail.com)

А.Е. Чекмарев - РУТ (МИИТ) (email: a.chekmarev@rut.digital)

данных: информация от валидаторов (платежных терминалов, которые используются для оплаты проезда) и данные телекоммуникационных операторов.

В первом случае - это, технически, некоторый лог, каждая строка которого содержит запись об одном проходе. Такая запись включает (это типичный CSV-файл) время прохода, идентификационную информацию валидатора, идентификационную информацию документа, использованного при проходе и тип этого документа (например, транспортная карта и ее номер). Оператор транспортной инфраструктуры может собирать такие данные по всем валидаторам (например, по всем валидаторам станции метро, по всем автобусам и т.д.). При этом возможны различные варианты предоставления таких данных. Например, идентификационная информация проездного документа может заменяться некоторым хэш-кодом для сохранения приватности или вообще отсутствовать. Схемы сохранения приватности здесь, на самом деле, такие же, как и при работе с данными телекоммуникационных операторов, где идентификация мобильного устройства (IMEI) также заменяется абстрактным хэшем. Технически эти данные собираются как журнал для платежной системы на транспорте, и по этой причине сбор такой информации, на самом деле, необходим транспортной системе. В целом, эта технология давно отработана и используется, например, при планировании развития транспортных систем в городе.

При этом, технически, проездной документ может предъявляться как однократно (при входе), так и дважды. В терминах социальных сетей - это события check-in и check-out. В московском регионе, данными которого мы оперируем, используются оба подхода. Метро работает исходя из однократного применения (использования) проездного документа, городские железные дороги, которые интегрированы с метро, требуют предъявлять проездной документ дважды - на входе и на выходе.

Если проездной документ предъявляется дважды, то в логах валидаторов будут две строки с одинаковым хэш-кодом проездного документа и таким образом можно непосредственно установить начальную и конечную точки маршрута, а также время поездки.

Если проездной документ предъявляется однократно, то для определения маршрута используются эвристические методы. Самое простое (и наиболее часто используемое) предположение состоит в том, что, если у нас есть информация о проходе на станции A по документу D в момент времени t :

$$\langle t, A, D \rangle$$

то следующее по времени использование того же проездного документа

$$\langle t_1, B, D \rangle \text{ где } t_1 > t$$

как раз и определяет конечную станцию B маршрута, начавшегося в A

Интерпретация достаточно прозрачная: станция, где проездной документ используется следующий раз - это

как раз та станция, до которой пассажир и доехал. С достаточно большой точностью, такая простая эвристика позволяет восстанавливать маршруты.

Анализируя время начала (окончания) маршрутов можно сделать предположения о месте проживания (станции, где маршруты завершаются вечером и начинаются утром), а также месте работы (станции, где завершаются утренние маршруты) - все это так называемые предпочтительные маршруты, а также о характере занятости (отсутствие предпочтительных маршрутов).

Другая форма предоставления транспортной информации - это автоматическое построение так называемых матриц корреспонденции. В англоязычной литературе - OD (origin-destination) матрицы. Такие матрицы показывают количество пассажиров, отправившихся за некоторую единицу времени с одной станции до другой. Поддержка мобильной связи в метро означает наличие там базовых станций операторов [7]. А это, в свою очередь, означает, что оператор может фиксировать моменты, когда его абонент (мобильное устройство) переключился на подземную станцию (то есть - вошел в метро) и наоборот, когда переключился с подземной станции на наземную (то есть - вышел из метро). Исходя из этого, оператор знает начальную и конечную точку любой поездки. Далее эти данные могут быть агрегированы по времени, так, чтобы исключалась возможность отслеживания единственной поездки, и результатом и будет матрица корреспонденций. Эта матрица привязана к определенному временному интервалу (периоду, за который аккумулировались данные) и показывает количество пассажиров, которые перемещались от одной станции метро до другой (рис. 1).

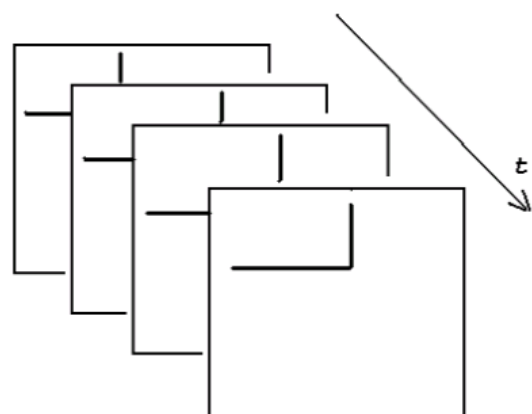


Рис. 1. Матрицы корреспонденции

Отметим, что в такой формулировке в качестве начальной и конечной "станции" могут выступать произвольные местоположения. Например, такие матрицы могут строиться просто для географических квадратов, определенного размера и показывать

количество мобильных абонентов, перемещающихся из одного квадрата в другой. Такая информация используется при планировании территорий [8]. В нашем случае, как отмечалось выше, используются данные именно по станциям метро.

III. АНАЛИЗ ДАННЫХ

Основная идея нашего анализа данных состоит в следующем. Мы хотим проверить гипотезу, что существуют (сформировались) районы (станции), где сосредоточен большой процент маломобильных граждан, пользующихся социальной поддержкой. Переводя эту гипотезу в анализ доступных нам данных – поиск районов (станций), где большое количество разовых поездок, и для оплаты проезда в большой степени используются льготные инструменты оплаты для специальных социальных групп.

Идея такого поиска возникла из ранних задач анализа использования станций городской железной дороги. В серии работ исследовались шаблоны использования станций железной дороги, где под шаблоном понималось распределение входов и выходов по времени [9, 10]. Идея состояла в проверке гипотезы о том, что есть некоторые сложившиеся шаблоны использования (например – пики по входам утром, когда пассажиры едут в Москву на работу, провал в использовании днем, и пики по выходу вечером, когда все возвращаются с работы), которые остаются стабильными по времени. Такие шаблоны были найдены, они сводились не только к традиционной маятниковой миграции. В частности, было показано, что есть станции, режим использования которых не был никак связан с рабочим графиком, чаще всего они использовались по выходным дням, а не рабочим. Единственное объяснение, которое как-то подходило для данной ситуации – около найденных станций массово располагались дачные поселения, где круглогодично жили, в основном, пенсионеры. Они периодически посещали город, но их трафик никак не был связан с рабочим графиком (рис. 2).

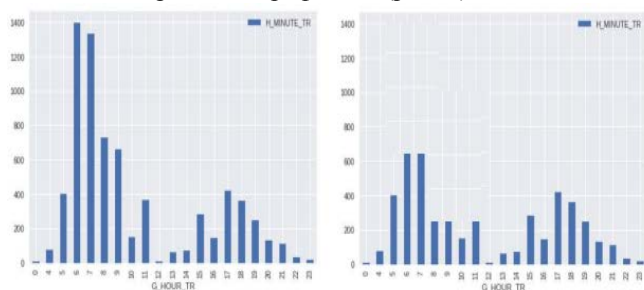


Рис. 2. Рабочий трафик (слева) и нерабочий трафик (справа)

Такая информация была важна, например, при проектировании интервального движения по городским железным дорогам. Очевидно, что для такого рода станций улучшение параметров движения ничего не

дает.

Отсюда и возникла идея попробовать такой же подход в анализе станций метро. Большая доля однократных поездок должна показывать, что люди редко покидают свой район, а использование льготных проездных документов является оценкой социального состава.

Для определения льготных поездок из массива входных данных отбирались соответствующие документы оплаты (карта Москвича и др.). В качестве индикатора разовых поездок использовались проездные документы на 1-2 поездки. Для того, чтобы выровнять условия по станциям разных размеров, использовались оценки долей соответствующих проходов в общем числе входов по станциям.

Далее мы следовали следующему алгоритму: доли социальных и разовых поездок считались для каждого дня выбранного месяца как в целом по городу, так и по отдельным станциям.

Если для выбранного месяца P – это распределение социальных поездок по всему городу, P_i – такое же распределение для i -й станции, то используя непараметрические критерии сравнения распределений (тест Колмогорова-Смирнова), можно отобрать множество станций $S = \{S_i\}$, где статистика с социальными проходами отличается от общегородской. Аналогичным образом может быть получено другое множество станций S^* , где статистика входов с одно или двукратными поездками отличается от общегородской. Пересечение $S \cap S^*$ и даст набор станций с аномальной статистикой использований социальных документов и однократными поездками. Такие станции были найдены для трех исследованных в вычислительном эксперименте месяцев.

IV. ЗАКЛЮЧЕНИЕ

В работе представлена модель оценки социального деления в городе, основанная на транспортных данных. Показано, что в современных условиях информация о поездках может служить базисом для получения данных о сложившихся социальных кластерах в городе. Это еще одно свидетельство в пользу того, что транспортные данные могут использоваться как измеритель городских процессов. Более того, непрерывный характер таких данных позволяет оценивать не только статические состояния, но и, что наиболее ценно – динамику процессов. В работе показано получение картины текущих кластеров. Периодические расчеты по представленным алгоритмам позволят оценить возможные изменения, которые будут отражением, как демографических процессов, так и естественным развитием города, миграцией населения и т.д. Важным моментом с точки зрения городских служб является то, что для анализа используется собираемая в обязательном порядке информация об оплате поездок. Это означает, что от городских служб не требуется

организации какой-то специальной схемы (сети) сбора данных. Переиспользование существующих данных (каналов сбора данных) делает такой анализ экономически выгодным.

БЛАГОДАРНОСТИ

Мы благодарны сотрудникам Научно-образовательного центра "Цифровые высокоскоростные транспортные системы" Российской открытой академии транспорта за ценные обсуждения данной работы.

БИБЛИОГРАФИЯ

- [1] Чешкова, Анастасия. "Методологические подходы к изучению городской пространственной сегрегации." *Российское городское пространство: попытка осмысления* (2000): 13-38.
- [2] Волков, А. А. О задачах создания эффективной инфраструктуры среды обитания / А. А. Волков, Д. Е. Намиот, М. А. Шнепс-Шнеппе // *International Journal of Open Information Technologies*. – 2013. – Т. 1. – № 7. – С. 1-10.
- [3] Новая пятилетка ВІМ - инфраструктура и умные города / В. П. Куприяновский, С. А. Сиягов, Д. Е. Намиот [и др.] // *International Journal of Open Information Technologies*. – 2016. – Т. 4. – № 8. – С. 20-35.
- [4] Куприяновский, В. П., et al. "Интеллектуальная мобильность и мобильность как услуга в Умных Городах." *International Journal of Open Information Technologies* 5.12 (2017).
- [5] Куприяновский, В. П., et al. "Цифровая трансформация экономики, железных дорог и умных городов. Планы и опыт Великобритании." *International Journal of Open Information Technologies* 4.10 (2016).
- [6] Namiot, Dmitry, and Oleg Pokusaev. "On mobility patterns in Smart City." *CEUR Workshop Proceedings*. 2019.
- [7] Kupriyanovsky, Vasily, et al. "On internet of digital railway." *International journal of open information technologies* 4.12 (2016): 53-68.
- [8] Куприяновский В. П. и др. Экономические выгоды применения комбинированных моделей ВІМ-ГИС в строительной отрасли. Обзор состояния в мире // *International Journal of Open Information Technologies*. – 2016. – Т. 4. – №. 5.
- [9] 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.
- [10] Namiot D., Pokusaev O., Kupriyanovsky V. Data Mining on the Use of Railway Stations // *AIST (Supplement)*. – 2018. – С. 66-74.

Social clusters in the city

Oleg Pokusaev, Dmitry Namiot, Alexander Chekmarev

Abstract— Digital methods of recording trips in transport today allow a quite accurate measurement of transport behavior. Digital payment tools and the development of telecommunications make it possible, in particular, to have complete information about passenger trips in the city. Accordingly, data on the movements of passengers become an indicator of the processes that take place in the city. In fact, transport data act as sensors, the measurements of which reflect the processes in the city system. For example, the distribution of entrances and exits of a particular metro station reflects the mode of use of the adjacent territory. In this paper, we consider the problem of assessing social clusters in a city based on information about the use of metro stations. Information on the form of payment for trips is used as the basis for such clustering.

Keywords— metro, travel payment, transport data

REFERENCES

- [1] Cheshkova, Anastasija. "Metodologicheskie podhody k izucheniju gorodskoj prostranstvennoj segregacii." Rossijskoe gorodskoe prostranstvo: popytka osmyslenija (2000): 13-38.
- [2] Volkov, A. A. O zadachah sozdaniya jeffektivnoj infrastruktury sredy obitaniya / A. A. Volkov, D. E. Namiot, M. A. Shneps-Shneppe // International Journal of Open Information Technologies. – 2013. – T. 1. – # 7. – S. 1-10.
- [3] Novaja pjatiletka BIM - infrastruktura i umnye goroda / V. P. Kuprijanovskij, S. A. Sinjagov, D. E. Namiot [i dr.] // International Journal of Open Information Technologies. – 2016. – T. 4. – # 8. – S. 20-35.
- [4] Kuprijanovskij, V. P., et al. "Intellektual'naja mobil'nost' i mobil'nost' kak usluga v Umnyh Gorodah." International Journal of Open Information Technologies 5.12 (2017).
- [5] Kuprijanovskij, V. P., et al. "Cifrovaja transformacija jekonomiki, zheleznyh dorog i umnyh gorodov. Plany i opyt Velikobritanii." International Journal of Open Information Technologies 4.10 (2016).
- [6] Namiot, Dmitry, and Oleg Pokusaev. "On mobility patterns in Smart City." CEUR Workshop Proceedings. 2019.
- [7] Kuprijanovskij, Vasily, et al. "On internet of digital railway." International journal of open information technologies 4.12 (2016): 53-68.
- [8] Kuprijanovskij V. P. i dr. Jekonomicheskie vygody primeneniya kombinirovannyh modelej BIM-GIS v stroitel'noj otrasli. Obzor sostojanija v mire //International Journal of Open Information Technologies. – 2016. – T. 4. – #. 5.
- [9] 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.
- [10] Namiot D., Pokusaev O., Kuprijanovskij V. Data Mining on the Use of Railway Stations //AIST (Supplement). – 2018. – S. 66-74.