

Визуальный анализ данных пассажиропотоков московского метрополитена

А.В. Маханкин, Д.Е. Намиот

Аннотация— Целью данной работы является создание четкой визуализации пассажиропотока для системы метрополитена г. Москвы. В работе проводится предварительная подготовка метаданных с входных групп турникетов, расположенных на станциях московского метрополитена и на их основе выполняется визуальный анализ пассажиропотока. Исследование было основано на информации за первый квартал 2020 года. Визуализация была создана с использованием различных методов и программного обеспечения: Система управления базами данных (СУБД) в которой хранится первичная информация для анализа— PostgreSQL, инструмент визуализации - Kepler.gl, фреймворк для языка программирования Java - Spring Framework, был использован для создания гибкого обращения к СУБД и получения результатов для Kepler.gl. Кроме того, рассматривается краткий обзор по использованию выбранного инструмента визуализации и его технических возможностей. На основе полученных результатов визуального анализа удалось определить базовые шаблоны перемещения пассажиров. С практической точки зрения, результаты помогают на раннем этапе выявить аномалии в движении и предпринять меры. Поэтому данный подход является важным, если не критическим, элементом интеллектуальных транспортных систем (ИТС) и систем умного города. Результаты проделанной работы представляют интерес для специалистов в области градостроительства и урбанистики.

Ключевые слова— визуальный анализ данных, анализ пассажиропотока, общественный транспорт, система метрополитена г. Москвы.

I. ВВЕДЕНИЕ

Развитие общественного транспорта является одной из важнейших целей городского планирования. Автомобильное движение занимает слишком много городского пространства [1]. Проектирование дорожной системы, ориентированной на автомобили, приводит только к увеличению трафика, увеличению пробок и, как следствие, к большему количеству аварий. Большинство современных городских проектов направлены на уменьшение количества автомобилей и закрытие определенных районов города от движения транспорта [2]. Однако для успешного создания нового городского типа необходимо иметь адекватное сочетание наземного и подземного общественного

транспорта. Развитие системы общественного транспорта приводит к снижению автомобильного трафика, помогает сделать городскую среду намного более комфортной для его жителей и предоставляет возможности не только для транспортной мобильности, но и для социальной мобильности [2].

В России вопрос баланса между автомобильными дорогами и системой общественного транспорта является более сложным, чем в других странах. Крупные городские агломерации имеют высокий уровень урбанизации, что сочетается с плохим финансированием транспортной системы. Только Московская агломерация может позволить себе реализовать крупные инфраструктурные проекты, такие как новые городские железнодорожные системы или новая кольцевая линия метро. За последние десять лет в несколько раз выросли темпы строительства новых линий и станций метро, заработали новые системы наземного железнодорожного сообщения, появились новые дороги и транспортные развязки. В столице открылось 116 км линий и 56 станций метро [3].

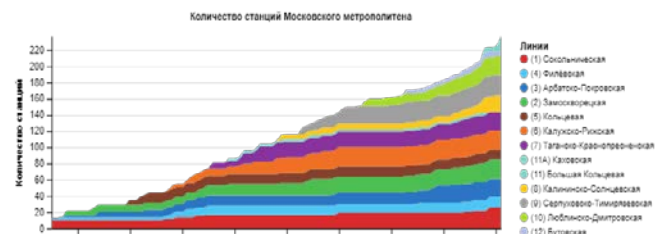


Рис. 1. Рост количества станций Московского метрополитена. Источник

Системы железных дорог и метро в больших городах часто страдают от задержек и аварий, и крайне важно отслеживать их влияние на потоки людей для управления поездами и распределения персонала. Такие мероприятия, как концерты и футбольные матчи, собирают десятки тысяч людей на ближайших станциях [3]. Количественное понимание таких изменений важно для оптимизации обычных операций и подготовки к следующим общественным мероприятиям. Визуализация изменений в потоках людей в широком диапазоне пространственно-временного пространства позволяет нам исследовать влияние таких явлений. В статье [4] проанализировали структуру загруженности станций метро в Лондоне. В [5][25] проанализировали пространственно-временную плотность пассажиров для одной линии метро в Сингапуре. В другой работе [6] показали полезность 3D-полос траекторий для визуализации данных атрибутов траектории.

Обычно основными данными для развития городской транспортной системы являются объемы

Статья получена 11 мая 2022.

Работа представляет собой результаты магистерской диссертации.

А. В. Маханкин, МГУ имени М. В. Ломоносова (e-mail: makhankin73@gmail.com).

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

пассажиропотока на существующих станциях. Однако семантические данные не подходят для восприятия людьми без определенных технических навыков [7]. Как пример - общественные организации. Поэтому необходимо визуализировать пассажиропоток понятным способом, чтобы сделать его полезным для общественных обсуждений и будущего городского планирования [1]. В настоящее время развитие сбора данных о дорожном движении приводят к росту проектов визуализации трафика. Они полезны для анализа транспортных сетей. В Китае существуют системы визуальной аналитики, основанные на данных смарт-карт и социальных сетей [8]. Данные о дорожном движении представлены в виде многослойной 3D-визуализации, основанной на пространственно-временных предпочтениях пассажиров [9]. Существует также визуализация данных смарт-карт и активности в Twitter для Токийского метрополитена [10][24]. Н. Андриенко и Г. Андриенко работают над большим количеством проектов визуализации данных, часть из которых связана с пассажирскими перевозками. Они связаны с личной активностью и инициативой SoBigData [11], пространственной связью с общественным транспортом [12] визуализацией графиков транспортной сети. Визуализация данных происхождения и назначения (OD) и моделирование пассажиропотока также очень популярны [13].

К сожалению, в приведенных выше работах нет никакой связи между визуализацией дорожного движения и картой города, что делает невозможным пространственное исследование [14].

Поэтому данная статья представляет собой одну из первых работ, посвященных визуальному анализу транспортных потоков метрополитена города Москвы.

II. ПРОБЛЕМЫ ПАССАЖИРОПОТОКОВ СИСТЕМЫ МЕТРОПОЛИТЕНА

В настоящее время, с увеличением населения мегаполисов, очевидно, растёт нагрузка и на их транспортные системы. В любой из них принципиально важной задачей является наблюдение и регулирование транспортных потоков, то есть, перемещений внутри этой системы.

Пассажиропоток Московского метрополитена является одним из самых высоких в мире. По количеству пассажиров, перевозимых в год, он уступает только Пекинскому, Токийскому, Шанхайскому, Сеульскому и метрополитену Гуанчжоу [15].



Рис. 2. Очередь на станцию «Выхино» в утренние часы. Источник: [16]

Доля метрополитена в перевозке пассажиров в Москве составляет 48 %.

К числу проблем, которые существуют у московского метрополитена, можно отнести:

- Перегрузки на пересадочных станциях с пригородного железнодорожного сообщения и автобусных станциях (ТПУ) в утренний и вечерний час пик;
- Возникновение большого скопления людей в вестибюлях, при этом увеличиваются риски несчастных случаев на станции;



Рис. 3. Скопление людей при задержке прихода поездов на станцию. Источник: [17]

- Высокая нагрузка на отдельные станции радиальных веток, по сравнению с другими ветками и станциями на них;

В рамках настоящей статьи наибольший интерес представляют такие проблемы, как возникновение больших очередей на станциях и их переполнение. Для решения данных проблем используется несколько подходов, в том числе расчет показателей пассажиропотока, моделирование (прогнозирование) пассажиропотока, визуальный анализ данных и другие подходы.

- Для количественной оценки пассажиропотока могут использоваться различные показатели, такие как: количество прошедших пассажиров через вестибюли станции метро;
- Средняя длина пути пассажиров
- Расстояние между соседними станциями
- Среднее время ожидания подвижного состава метро и др.

III. ИСХОДНЫЕ ДАННЫЕ

Предоставленный для исследования набор данных представляет из себя файл CSV формата, содержащую информацию о передвижениях пассажиров Московского метрополитена в период с 1 февраля по 1 марта 2020 года. Объем файла 10 Гб, количество записей - 216 149 734 строк. Каждая строка содержит следующую информацию:

- Дата/время валидации билета;
- ID вестибюля;
- Номер валидатора;
- Номер билета;
- Номер кристалла;
- ID тип билета;
- Тип прохода;

- Порядковый номер поездки по билету (если будет проставлен в системе);
- Количество оставшихся поездок по билету (если будет проставлено в системе)

Фрагмент исходных данных показан на рис. 3

```

1 "date","vestibul_id","validator_id","ticket_num","cristal_num","ticket_type","passage_type","ticket_order_num","ticket_remains_count"
2 "2020-03-01 04:59:42+00",137,1194,102033919,10204064944912,4112,0,491,0
3 "2020-03-01 04:59:43+00",137,1204,102033919,10204064944912,4112,0,492,0
4 "2020-03-01 04:59:43+00",137,10935,102033919,10204064944912,4112,0,1099,0
5 "2020-03-01 04:59:44+00",2239,14385,1270024349,11493404790032,4112,0,24,0
6 "2020-03-01 04:59:45+00",2020,12179,90033912,4293195,4112,0,2205,0
    
```

Рис. 4. Фрагмент исходного датасета. Источник: получен автором.

IV. ПРЕДОБРАБОТКА ДАННЫХ И ИНСТРУМЕНТ ВИЗУАЛИЗАЦИИ

Важно отметить, что визуализация данных осуществляется в несколько этапов: преобразование данных, визуальное отображение и преобразование вида.

В рамках настоящего исследования наибольший интерес представляет именно визуализация пассажиропотоков.

Далее, будут рассмотрены основные моменты перед непосредственной визуализацией исходных данных.

A. Предобработка данных

Работать сразу с таким файлом не представляется возможным. Большинство популярных текстовых редакторов просто не способны даже открыть файл. Под данную задачу отлично подходит СУБД [20]:

- Можно хранить большое количество информации
- Информацию можно разделить на мелкие сущности
- Легко проводить аналитические запросы с помощью языка SQL
- Результат можно экспортировать в файл для последующей загрузки в инструменты визуализации

Для начала подготовим схему БД для загрузки в нее исходных данных. Итоговая схема базы данных представлена на рис. 5

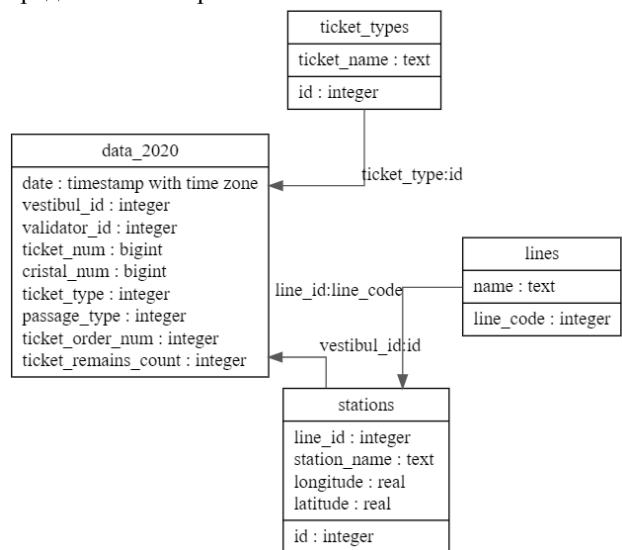


Рис. 5. Схема базы данных. Источник: составлено автором по результатам работы.

Связи между полями в таблицах отображены стрелками (каждая стрелка имеет комментарий – имя колонок, между которыми установлено отношение). Описание схемы приведено в нижеследующей таблице (табл. 1)

Таблица 1. Описание схемы СУБД

Название	Описание
data_2020	Данные с турникетов станций метро
stations	Информация о станциях метро и их географические координаты
lines	Линии московского метрополитена
ticket_types	Перечень типов входных билетов

Источник: составлено автором на основе полученных результатов работы.

Информация о географическом расположении станций метрополитена была взята из открытых источников [22],[23].

B. Преобразование данных для Kepler.gl

Для удобства получения данных по нужным нам параметрам фильтрации был реализован REST API сервис на основе фреймворка для языка программирования Java – Spring Framework. Так как инструмент визуализации Kepler.gl поддерживает ввод данных из внешней ресурса с использованием HTTP протокола, поэтому с использованием фреймворка создадим несколько REST контроллеров [21], которые будут выдавать результаты из базы данных в формате JSON, поддерживаемый Kepler.gl [19].

Адрес и описание контроллеров приведено в табл. 2

Таблица 2 Описание контроллеров для получения данных из базы данных.

Адрес	Входные аргументы	Описание контроллера
http://localhost:8080/api/getPassengerCountBetweenDateByStations?startDate=<1>&&endDate=<2>	<1> - дата, с которой требуется получить информацию, <2> - конечная дата.	Получение данных о кол-ве пассажиров на каждой станции метрополитена в заданном диапазоне дат
http://localhost:8080/api/getPassengerCountBetweenDateByStationsAndTicketType?startDate=<1>&&endDate=<2>&&ticketType=<3>	<1> - дата, с которой требуется получить информацию, <2> - конечная дата, <3> - тип билета (описание типов представлено в таблице БД – ticket_types)	Получение данных о кол-ве пассажиров по типу билета в диапазоне дат

<code>http://localhost:8080/api/getPassengerCountBetweenDateByStationsAndTicketTypes?startDate=<1>&&endDate=<2></code>	<1> - дата, с которой требуется получить информацию, <2> - конечная дата,	Получение данных о кол-ве пассажиров по каждому типу билета в диапазоне дат
<code>http://localhost:8080/api/getPassengerCountBetweenDateByStationsWithTimeStep?startDate=<1>&&endDate=<2>&&line<3>&&step=<4></code>	<1> - дата, с которой требуется получить информацию, <2> - конечная дата, <3> - код линии метрополитена, <4> - временной шаг (в мин), за который надо сгруппировать количество проходов на станции.	Получение данных о входах на станциях с заданной линии метро за заданный временной период и временной шаг. По умолчанию, за каждые 5 минут.

Источник: составлено автором на основе полученных результатов работы.

В результате мы получили гибкие инструменты, не требующие знания SQL языка для получения данных, а значит данным программным комплексом может воспользоваться рядовой пользователь.

С. Инструмент визуализации Kepler.gl

Все необходимые процедуры подготовки данных были выполнены, теперь мы можем заняться визуализацией.

В настоящее время для визуального анализа данных существует большое число различных инструментов, среди которых приложения Kepler.gl, Carto и Locale.ai. Данные инструменты широко используются крупными компаниями в различных целях. Так, самым известным примером визуализации данных при помощи Kepler.gl является компания Uber Engineering [19], которая широко использовала этот инструмент для анализа дорожного движения в различных городах мира (к примеру, анализ опасных



Рис. 6. Интерфейс инструмента визуализации Kepler.gl. Источник: [17]

участков дорог в Манхэттене) [10][11]

В данной работе выбор был остановлен на Kepler.gl. Он представляет собой высокопроизводительное веб-приложение для визуализации больших гео-данных и может использоваться в большом числе различных направлений анализа и визуализации данных.

V. ВИЗУАЛИЗАЦИЯ ДАННЫХ ПАССАЖИРОПОТОКОВ ПРИ ПОМОЩИ KEPLER.GL

В исходных данных представлено большое количество аргументов, но стоит выделить наиболее важные для построения визуальных моделей:

- Дата и время прохода через турникет;
- Код вестибюля;
- Код линии метрополитена (не был представлен в датасете);
- Тип билета

Так как в исходном наборе данных было только указание на вестибюль прохода, то при подготовке данных автором была подготовлена связь вестибюлей со станцией, а станции с линией метрополитена. Это было сделано для того, чтобы можно было не просто визуализировать данные точечной станции, а сравнивать различные ветки между собой в произвольный промежуток времени.

Это позволяет определять наиболее загруженные пассажирами станции и, принимая точно решения об управлении пассажиропотоком на них.

Московский метрополитен имеет большое количество линий и станций на них, поэтому для простоты в статье будет рассмотрена визуализация одной из самой загруженной ветки – «Таганско-Краснопресненская». И попробуем найти различия в поведении пассажиропотока в будние дни и в выходные, так же объяснить причины отклонения от будничного паттерна.

Для визуализации входа на станции был выбран столбчатый слой, где каждый столбец — это станция, а высота столбца и цвет — зависят от количество валидированных билетов (чем выше столбец (темнее цвет), тем больше входящий поток на станцию). Как

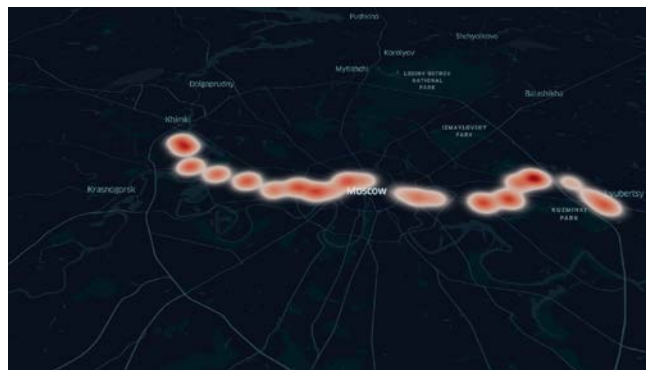


Рис. 7. Пример отображения данных с использованием слоя Heatmap. Источник: составлено автором

альтернативу столбчатому слою, под нашу задачу, можно рассмотреть Heatmap слой (рис. 7), на котором интенсивность данных представляется в географических точках при помощи разных цветов. Чем ярче цвет, тем

«горячее» станция, то есть количество входов сильно больше, чем на соседних [20].

Так же количество проходов было сгруппировано по временным интервалам – каждые 5 минут. Это сделано для того, чтобы работал интервальный фильтр, и можно было динамически наблюдать изменение картины движения.

На рис. 8 представлена столбчатая визуализация проходов на станции метро в период времени с 20:56-11:06 на дату 3 февраля 2020 года. Текущее время не попадает под будничные часы пик и это визуально наблюдается, что нет большого количества вхождений на станциях, расположенных в спальных районах. Пики наблюдаются только на входе у станций, которые находятся внутри третьего транспортного кольца.

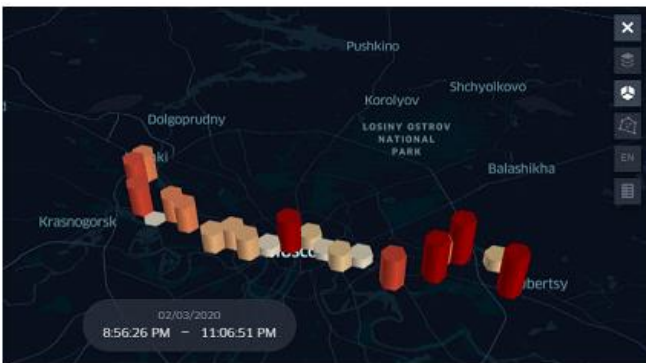


Рис. 8. Визуализация данных Таганско-Краснопресненской линии в будний день. Источник: составлено автором.

На рис. 9 представлена визуализация с аналогичным форматом, но на период времени с 20:56-23:06 на дату 8 февраля 2020 года. Здесь мы отчетливо видим нетипичные пики в районе станций, расположенных вблизи ТПУ или в густонаселенных спальных районах.

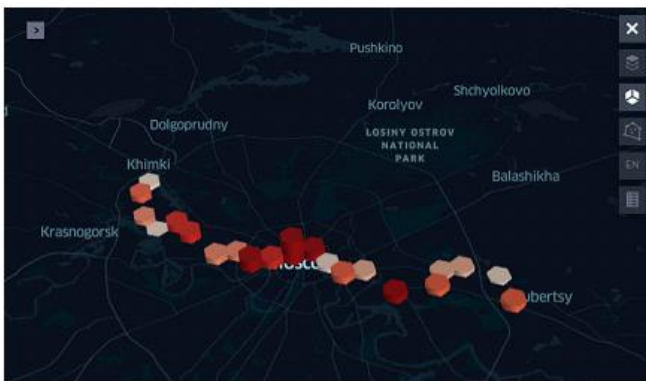


Рис. 9. Визуализация данных Таганско-Краснопресненской линии в выходной день. Источник: составлено автором.

Если пронаблюдать динамику во времени, то обнаружим, что люди едут в центральную часть города, где расположено большое количество общественно-развлекательных мест. Как пример, станция «Кузнецкий мост».

Так же, для подтверждения корректности визуализации, рассмотрим графики зависимости между временем и количеством входов.

На рис. 10 график за будний день и на нем четко видно основной паттерн передвижения людей –

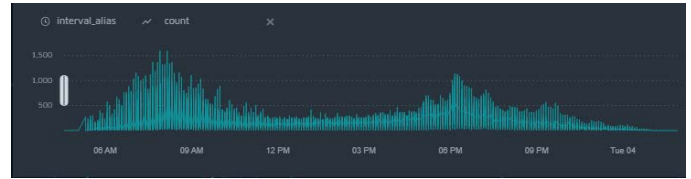


Рис. 10. График зависимости количества проходов в конкретный момент времени (будний день). Источник: составлено автором.

наблюдаем 2 пика:

- В утреннее время с 7 до 9 утра. В данный промежуток времени все устремляются на работу – с пересадочных станций пригородного сообщения, с автобусных станций, соединяющих Московскую область с городом, а также из спальных районов самой столицы.
- В вечернее время с 17 до 20 вечера. В данный промежуток времени картина полностью инвертируется. Люди из центральной части Москвы двигаются к ее окраинам или уезжают в область.

На рис. 11 график представлен за выходной день. Картина по плотности передвижения достаточно сильно различается. Видно, что пассажиропоток сильно упал и он более однородный. Но больший интерес представляет всплеск входов на станциях с 21 до 23

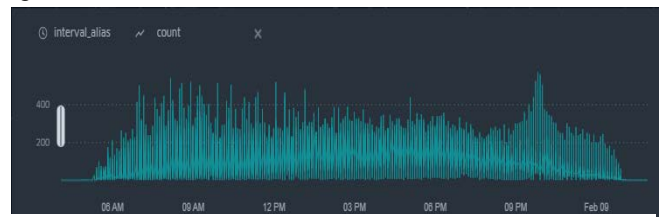


Рис. 11. График зависимости количества проходов в конкретный момент времени (выходной день). Источник: составлено автором.

ночи.

На основе описанных выше визуальных представлений данных с турникетов на станциях «Таганско-Краснопресненской» линии, можно сделать два вывода. Во-первых, видно, что паттерн и плотность передвижения потока в будний день отличается от движения в выходной день. Во-вторых, наблюдаются нетипичные пики на вход в вечернее время выходного дня на станциях, расположенных вблизи транспортно-пересадочных узлов или густонаселенных спальных районов. Наблюдение в динамике по времени, показывается, что после 23 ночи начинается большой отток людей со станций внутри третьего транспортного кольца, т. е. там, где находится большое количество общественно-развлекательных мест. В рассмотренный выходной день – самая загруженная станция оказалась «Кузнецкий мост».

VI. ОБ АНАЛИЗЕ ТРАНСПОРТНЫХ ДАННЫХ

Необходимо отметить, что анализ транспортных данных и, в частности, использование Kepler.gl имеет

уже некоторую историю в рамках лаборатории ОИТ факультета ВМК.

Начиная с работ, инициированных В.П. Куприяновским [27, 28], которые продвигали комплексный подход к анализу [29], до более практических работ, включая и магистерские диссертации, затрагивавших, в том числе, и визуальный анализ транспортных данных [30, 31, 32]. Это были, в частности, и пионерские работы в плане использования Kepler.gl.

Текущие направления работ связаны с разработкой и использованием цифровых двойников для транспортных систем, где визуализация занимает, естественно, не последнее место.

В связи с этими работами необходимо также отметить с благодарностью сотрудников Центра Высокоскоростных транспортных систем Российского университета транспорта [33] за их помощь в получении данных для анализа, а также оценке и рецензировании работ.

VII. ЗАКЛЮЧЕНИЕ

В настоящее время визуальный анализ данных является одним из наиболее перспективных и востребованных методов оценки большого объема данных о пассажиропотоке столичного метрополитена.

Актуальность такого метода с каждым годом все больше возрастает, что обусловлено развитием цифровизации городской инфраструктуры, быстрыми темпами строительства, как жилых кварталов, так и станций и линий метро, которые критически необходимы с ростом населения в городе.

В данной работе был использован один из наиболее продвинутых инструментов визуализации данных - Kepler.gl. Он идеально подошел для решения поставленной задачи. В сравнении с другими инструментами он имеет такие значимые преимущества, как открытый исходный код и большое число технических возможностей.

Именно при помощи Kepler.gl была осуществлена визуализация данных пассажиропотока на радиальной ветке Таганско-Краснопресненской линии московского метрополитена. Использовались данные с турникетов, расположенные на станциях.

В результате приведенного анализа были сделано несколько важных выводов. Во-первых, в период рабочей недели четко наблюдаются утренние и вечерние пики. Во-вторых, в выходные дни наблюдается нетипичное поведение в сравнении с будним шаблоном распределения людей. Происходит увеличение количества людей, которые входят на станции расположенные либо в спальных районах, либо в местах пересадок с пригородного транспорта. В-третьих, в выходные дни после 23.00 наблюдается высокая активность на станциях, внутри третьего транспортного кольца, что означает отток с мест, где очень плотно расположены общественно-развлекательные места и сооружения.

Качественные выводы, полученные в результате визуального анализа данных о пассажиропотоке

позволяют принимать сотрудникам Московского метрополитена управленческие решения по оптимизации пассажиропотока во избежание таких проблем, как возникновение больших очередей на перроне станции во время часов пик, а также в период выходных и праздничных дней, когда требуется быстро распределить людей в небольшом промежутке времени.

БИБЛИОГРАФИЯ

- [1] Куприяновский В. П. и др. Правительство, промышленность, логистика, инновации и интеллектуальная мобильность в цифровой экономике //Современные информационные технологии и ИТ-образование. – 2017. – Т. 13. – №. 1. – С. 74-96
- [2] I. Capra, C. Smith, and L. Capra. Avoiding the Crowds: Understanding Tube Station Congestion Patterns from Trip Data. In Proc. Urb-Comp'12, pages 134–141, 2012
- [3] L. Sun, D.-H. Lee, A. Erath, and X. Huang. Using Smart Card Data to Extract Passenger's Spatio-temporal Density and Train's Trajectory of MRT System. In Proc. UrbComp'12, pages 142–148, 2012.
- [4] Алешко Р. А. и др. Разработка методики визуализации и обработки геопространственных данных //Научная визуализация. – 2015. – Т. 7. – №. 1.
- [5] Goodwin, P., & Noland, R. B. (2003). Building new roads really does create extra traffic: a response to Prakash et al. *Applied Economics*, 35(13), 1451–1457. <https://doi.org/10.1080/0003684032000089872>
- [6] Беляков С. Л., Белякова М. Л., Савельева М. Н. Адаптивная к изменению структуры базы данных визуализация пространственных данных //Приборы и системы. Управление, контроль, диагностика. – 2016. – №. 1. – С. 25-32.
- [7] Card M. Readings in information visualization: using vision to think. – Morgan Kaufmann, 1999.
- [8] Ding X. et al. Viptra: Visualization and interactive processing on big trajectory data //2018 19th IEEE International Conference on Mobile Data Management (MDM). – IEEE, 2018. – С. 290-291.
- [9] Gonçalves T., Afonso A. P., Martins B. Cartographic visualization of human trajectory data: Overview and analysis //Journal of Location Based Services. – 2015. – Т. 9. – №. 2. – С. 138-166.
- [10] Bumgardner, B. (2016). Mapping NYC subway traffic: an interactive. <http://bryanbumgardner.com/mapping-nyc-subwaytraffic-an-interactive>
- [11] Chong, S. M. (2015). NYC subway traffic. <http://piratefsh.github.io/mta-maps/public/> (дата обращения: 11.05.2022)
- [12] Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S., Kitsuregawa, M. (2014). Visual fusion of mega-city big data: An application to traffic and tweets data analysis of Metro passengers. In 2014 IEEE International Conference on BigData (Big Data) (pp. 431–440). IEEE. <https://doi.org/10.1109/BigData.2014.7004260>
- [13] Sobral T., Galvão T., Borges J. Visualization of urban mobility data from intelligent transportation systems //Sensors. – 2019. – Т. 19. – №. 2.
- [14] Thomas J. J. Illuminating the path: the research and development agenda for visual analytics. – IEEE Computer Society, 2005.
- [15] Zeng L. et al. A passenger flow control method for subway network based on network controllability //Discrete Dynamics in Nature and Society. – 2018.
- [16] Открытый банк фотографий; URL: <https://ru.depositphotos.com/stock-photos> (дата обращения: 11.05.2022)
- [17] РИА новости; URL: <https://ria.ru> (дата обращения: 11.05.2022)
- [18] Misharin A., Namiot D., Pokusaev O. On Passenger Flow Estimation for new Urban Railways //IOP Conference Series: Earth and Environmental Science. – IOP Publishing, 2018. – Т. 177. – №. 1. – С. 012012.
- [19] Официальный сайт Kepler.gl. URL: <https://kepler.gl/> (дата обращения: 11.05.2022)
- [20] Официальный сайт документации PostgreSQL, URL: <https://www.postgresql.org/docs/> (дата обращения: 11.05.2022)
- [21] Официальный сайт документации Spring Framework, URL: <https://docs.spring.io/spring-framework/docs/current/reference/html/> (дата обращения: 11.05.2022)
- [22] Портал открытых данных Правительства Москвы – станции Московского метрополитена; URL:

- <https://data.mos.ru/classifier/7704786030-stantsii-moskovskogo-metropolitena> (дата обращения: 11.05.2022)
- [23] Документация API HeadHunter; URL: <https://github.com/hhru/api/blob/master/docs/areas.md> (дата обращения: 11.05.2022)
- [24] Shin, H. (2020). Analysis of subway passenger flow for a smarter city: knowledge extraction from Seoul metro's 'Untraceable' big data. *IEEE Access*, 8, 69296–69310. <https://doi.org/10.1109/ACCESS.2020.2985734>
- [25] Xiao, F., & Yu, G. (2018). Impact of a new metro line: analysis of metro passenger flow and travel time based on smart card data. *Journal of Advanced Transportation*, 2018, 9247102. <https://doi.org/10.1155/2018/9247102>
- [26] Tanaka, K. (1950). The relief contour method of representing topography on maps. *Geographical Review*, 40(3), 444–456. <https://doi.org/10.2307/211219>
- [27] Куприяновский В. П. и др. Цифровая трансформация экономики, железных дорог и умных городов. Планы и опыт Великобритании //International Journal of Open Information Technologies. – 2016. – Т. 4. – №. 10. – С. 22-31.
- [28] Соколов И. А. и др. Прорывные инновационные технологии для инфраструктур. Евразийская цифровая железная дорога как основа логистического коридора нового Шелкового пути //International Journal of Open Information Technologies. – 2017. – Т. 5. – №. 9. – С. 102-118.
- [29] Куприяновский В. П. и др. Цифровая экономика и Интернет Вещей-преодоление силоса данных //International Journal of Open Information Technologies. – 2016. – Т. 4. – №. 8. – С. 36-42.
- [30] Куприяновская Ю. В. и др. Умный контейнер, умный порт, ВІМ, Интернет Вещей и блокчейн в цифровой системе мировой торговли //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 3. – С. 49-94.
- [31] Medvedenko S., Namiot D. Visual analysis of railway passenger traffic data //International Journal of Open Information Technologies. – 2021. – Т. 9. – №. 6. – С. 51-60.
- [32] Nekraplonna M., Namiot D. The Analysis of Trajectories in Moscow Subway. – 2021.
- [33] ЦВТС РУТ <https://cvts.rut.digital/> (дата обращения: 11.05.2022)

Visual analysis of passenger traffic data of the Moscow subway

Artem Makhankin, Dmitry Namiot

Abstract — The purpose of this work is to create a clear visualization of passenger traffic for the Moscow metro system. The paper provides preliminary preparation of metadata from the entrance groups of turnstiles located at Moscow metro stations and visual analysis of passenger traffic is performed on their basis. The study was based on information for the first quarter of 2020. Visualization was created using various methods and software: Database Management System (DBMS) in which primary information for analysis is stored – PostgreSQL, visualization tool - Kepler.gl, a framework for the Java programming language – Spring Framework, was used to create a flexible access to the DBMS and obtain results for Kepler.gl. In addition, a brief overview of the use of the selected visualization tool and its technical capabilities is considered. Based on the results of visual analysis, it was possible to determine the basic patterns of passenger movement. From a practical point of view, the results help to identify anomalies in movement at an early stage and act. Therefore, this approach is an important, if not critical, element of intelligent transport systems (ITS) and smart city systems. The results of the work done are of interest to specialists in the field of urban planning and urban studies.

Key words — visual data analysis, passenger traffic analysis, public transport, Moscow subway.

REFERENCES

- [1] Kupriyanovsky V. P. et al. Government, industry, logistics, innovation and intellectual mobility in the digital economy //Modern information technologies and IT education. – 2017. – Vol. 13. – No. 1. – pp. 74-96
- [2] I. Ceapa, C. Smith, and L. Capra. Avoiding the Crowds: Understanding Tube Station Congestion Patterns from Trip Data. In Proc. Urb-Comp'12, pages 134–141, 2012
- [3] L. Sun, D.-H. Lee, A. Erath, and X. Huang. Using Smart Card Data to Extract Passenger's Spatio-temporal Density and Train's Trajectory of MRT System. In Proc. UrbComp'12, pages 142–148, 2012.
- [4] Aleshko R. A. et al. Development of methods for visualization and processing of geospatial data //Scientific visualization. – 2015. – Vol. 7. – No. 1.
- [5] Goodwin, P., & Noland, R. B. (2003). Building new roads really does create extra traffic: a response to Prakash et al. Applied Economics, 35(13), 1451–1457. <https://doi.org/10.1080/0003684032000089872>
- [6] Belyakov S. L., Belyakova M. L., Savelyeva M. N. Spatial data visualization adaptive to changes in the database structure //Devices and systems. Management, control, diagnostics. - 2016. – No. 1. – pp. 25-32.
- [7] Card M. Readings in information visualization: using vision to think. – Morgan Kaufmann, 1999.
- [8] Ding X. et al. Viptra: Visualization and interactive processing on big trajectory data //2018 19th IEEE International Conference on Mobile Data Management (MDM). – IEEE, 2018. – C. 290-291.
- [9] Gonçalves T., Afonso A. P., Martins B. Cartographic visualization of human trajectory data: Overview and analysis //Journal of Location Based Services. – 2015. – T. 9. – №. 2. – C. 138-166.
- [10] Bumgardner, B. (2016). Mapping NYC subway traffic: an interactive. <http://bryantumgardner.com/mapping-nyc-subwaytraffic-an-interactive>
- [11] Chong, S. M. (2015). NYC subway traffic. <http://piratefsh.github.io/mta-maps/public/>
- [12] Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S., Kitsuregawa, M. (2014). Visual fusion of mega-city big data: An application to traffic and tweets data analysis of Metro passengers. In 2014 IEEE International Conference on BigData (Big Data) (pp. 431–440). IEEE. <https://doi.org/10.1109/BigData.2014.7004260>
- [13] Sobral T., Galvão T., Borges J. Visualization of urban mobility data from intelligent transportation systems //Sensors. – 2019. – T. 19. – №. 2.
- [14] Thomas J. J. Illuminating the path: the research and development agenda for visual analytics. – IEEE Computer Society, 2005.
- [15] Zeng L. et al. A passenger flow control method for subway network based on network controllability //Discrete Dynamics in Nature and Society. – 2018.
- [16] Open photo bank; URL: <https://ru.depositphotos.com/stock-photos>
- [17] RIA Novosti; URL: <https://ria.ru>
- [18] Misharin A., Namiot D., Pokusaev O. On Passenger Flow Estimation for new Urban Railways //IOP Conference Series: Earth and Environmental Science. – IOP Publishing, 2018. – T. 177. – №. 1. – C. 012012.
- [19] Official website Kepler.gl. URL: <https://kepler.gl/>
- [20] Official documentation website PostgreSQL, URL: <https://www.postgresql.org/docs/>
- [21] Official documentation website Spring Framework, URL: <https://docs.spring.io/spring-framework/docs/current/reference/html/>
- [22] Moscow Government Open Data Portal – Moscow Metro stations; URL: <https://data.mos.ru/classifier/7704786030-stantsii-moskovskogo-metropolitena>
- [23] Web API documentation HeadHunter; URL: <https://github.com/hhru/api/blob/master/docs/areas.md>
- [24] Shin, H. (2020). Analysis of subway passenger flow for a smarter city: knowledge extraction from Seoul metro's 'Untraceable' big data. IEEE Access, 8, 69296–69310. <https://doi.org/10.1109/ACCESS.2020.2985734>
- [25] Xiao, F., & Yu, G. (2018). Impact of a new metro line: analysis of metro passenger flow and travel time based on smart card data. Journal of Advanced Transportation, 2018, 9247102. <https://doi.org/10.1155/2018/9247102>
- [26] Tanaka, K. (1950). The relief contour method of representing topography on maps. Geographical Review, 40(3), 444–456. <https://doi.org/10.2307/211219>
- [27] Kuprijanovskij V. P. i dr. Cifrovaja transformacija jekonomiki, zheleznyh dorog i umnyh gorodov. Plany i opyt Velikobritanii //International Journal of Open Information Technologies. – 2016. – T. 4. – №. 10. – S. 22-31.
- [28] Sokolov I. A. i dr. Proryvnye innovacionnye tehnologii dlja infrastruktur. Evrazijskaja cifrovaja zheleznaia doroga kak osnova logisticheskogo koridora novogo Shelkovogo puti //International Journal of Open Information Technologies. – 2017. – T. 5. – №. 9. – S. 102-118.
- [29] Kuprijanovskij V. P. i dr. Cifrovaja jekonomika i Internet Veshhej-preodolenie silosa dannyh //International Journal of Open Information Technologies. – 2016. – T. 4. – №. 8. – S. 36-42.
- [30] Kuprijanovskaja Ju. V. i dr. Umnyj kontejner, umnyj port, BIM, Internet Veshhej i blokchejn v cifrovoj sisteme mirovoj trgovli //International Journal of Open Information Technologies. – 2018. – T. 6. – №. 3. – S. 49-94.
- [31] Medvedenko S., Namiot D. Visual analysis of railway passenger traffic data //International Journal of Open Information Technologies. – 2021. – T. 9. – №. 6. – S. 51-60.
- [32] Nekraplonna M., Namiot D. The Analysis of Trajectories in Moscow Subway. – 2021.
- [33] CVTS RUT <https://cvts.rut.digital/> (data obrashhenija: 11.05.2022)