

Визуальный анализ данных пассажиропотока железнодорожного транспорта

С.А. Медведенко, Д.Е. Намиот

Аннотация — В настоящей статье рассматриваются особенности визуального анализа данных пассажиропотока железнодорожного транспорта при помощи приложения для визуализации Kepler.gl. Прежде всего рассматриваются проблемы, связанные с ростом пассажиропотока городов, которые возникают вследствие увеличения численности населения, роста городов и развития транспортной инфраструктуры. В результате этого, обосновывается необходимость эффективного управления пассажиропотоком железнодорожного транспорта и решения названной проблемы, в том числе за счет визуального анализа данных. Далее проводится сравнительная характеристика существующих инструментов визуализации и доказываем, что наиболее продвинутым из них является приложение Kepler.gl. За счет открытого исходного кода и большого числа технических возможностей, он является самым продвинутым и совершенным инструментом для решения проблемы возникновения очередей и скопления большого числа пассажиров на железнодорожных станциях. С помощью приложения Kepler.gl были представлены данные, собранные с проходов через турникеты на железнодорожных станциях. На основе визуализации этих данных удалось прийти к выводу о том, что основные пассажиропотоки перемещаются между крупными городами, причем количество отправок из них превышает количество прибытий.

Ключевые слова — визуализация данных, визуальный анализ данных, визуализация пассажиропотока, железнодорожный транспорт

I. ВВЕДЕНИЕ

На сегодняшний день железнодорожный транспорт в городах является важным средством передвижения для большого количества людей. Ежедневно сотни тысяч или даже миллионы пассажиров пользуются таким транспортом для междугороднего перемещения по России на длинные расстояния, для пригородного

сообщения, а также для перемещения в пределах городов (например, в Москве действует несколько маршрутов пригородного сообщения, проходящие через город).

Кроме того, города и другие населенные пункты постоянно растут и развиваются, численность их населения возрастает ежегодно, а границы сдвигаются. Следовательно, возникает необходимость расширения транспортной инфраструктуры, в том числе и железнодорожной. С учетом такой интенсивности перемещения и ростом населенных пунктов могут возникать проблемы, такие как неравномерная загруженность станций. Поэтому, возникает острая необходимость эффективного распределения потоков пассажиров в целях избегания возможных проблем. Для этого принимаются решения, основанные на данных о пассажиропотоке на железнодорожных станциях.

Анализ и обработка данных пассажиропотоков на железнодорожном транспорте требует применения специальных методов и подходов. Это обусловлено тем, что такие данные содержат большое количество информации о направлениях следования пассажиров, типах используемых билетов (например, льготные, взрослые, детские), о дате и времени проходов пассажиров, направлении (вход или выход), станция отправления и прибытия пассажиров. Такие данные преимущественно собираются при помощи турникетов, расположенных на железнодорожных станциях, через которые проходят пассажиры для посадки на транспорт.

Настоящая статья написана по результатам магистерской диссертации, выполненной на факультете ВМК МГУ имени М.В. Ломоносова. Анализу транспортных потоков именно для железных дорог было посвящено достаточно много работ в лаборатории Открытых информационных технологий [27-29], включая как общие подходы к такому анализу, так и конкретные методы обработки.

В ряде случаев применение классических методов анализа больших гео-данных является недостаточным, а в ряде случаев они и вовсе не применимы. Поэтому визуальный анализ данных позволяет более эффективно представлять имеющиеся данные, объективно оценивать текущую ситуацию, а также принимать правильные управленческие решения. Данная статья представляет собой одну из первых работ, посвященных именно визуальному анализу транспортных потоков.

II. ЦИФРОВИЗАЦИЯ И ВИЗУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

В последние годы одной из наиболее актуальных тенденций развития всех сфер является цифровизация.

Статья получена 10 мая 2021.

Работа представляет собой результат магистерской диссертации.

С. А. Медведенко, МГУ имени М. В. Ломоносова

(e-mail: stepahardin@gmail.com).

Д. Е. Намиот, МГУ имени М. В. Ломоносова

(e-mail: dnamiot@gmail.com).

Она представляет собой использование и внедрение цифровых технологий в промышленность, экономику, образование, культуру, сферу услуг и пр. При этом, к числу цифровых технологий относятся Большие данные (англ. Big Data), Интернет вещей (англ. Internet of Things – IoT), технологии дополненной (англ. Augmented reality – AR) и виртуальной реальности (англ. Virtual reality – VR), блокчейн (англ. Blockchain), технологии распределенного реестра, сети 5G и многое другое [18]. Использование данных технологий в настоящее время находится на довольно высоком уровне и сейчас тяжело представить крупную компанию, которая для повышения эффективности своей деятельности не использовала бы цифровые технологии. Актуально это в том числе и в сфере визуального анализа данных пассажиропотоков железнодорожного транспорта.

С развитием цифровых технологий происходит развитие городов и их транспортной инфраструктуры. При этом города с развитой инфраструктурой и транспортной доступностью привлекают большее число людей для проживания и работы в них. Все это увеличивает пассажиропотоки между различными населенными пунктами, а как следствие, появляется все больше информации об их перемещении, которую необходимо собирать, обрабатывать и анализировать. Поэтому в условиях цифровизации возрастает актуальность принятия решений, основанных на данных.

III. ПРОБЛЕМЫ ПАССАЖИРОПОТОКОВ НА ЖЕЛЕЗНОДОРОЖНОМ ТРАНСПОРТЕ

С развитием городов и расширением транспортной инфраструктуры появляется ряд проблем в области железнодорожного транспорта, к числу которых можно отнести:

- возникновение больших очередей на железнодорожных станциях, их переполнение и создание заторов;
- большой поток пассажиров на направлениях пригородного сообщения;
- перегрузки железнодорожного транспорта во время часов пик;
- высокая загруженность определенных железнодорожных станций, в сравнении с другими станциями и другие проблемы.

В рамках настоящей статьи наибольший интерес представляют такие проблемы, как возникновение больших очередей на железнодорожных 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.

Так, для оценки пассажиропотока могут использоваться различные показатели. К их числу можно отнести следующие показатели, при помощи которых может оцениваться поток пассажиров между населенными пунктами [25]:

- количество пассажиров;
- коэффициент кластеризации;
- расстояние между транспортными узлами;
- средняя длина пути пассажиров и пр.

В свою очередь, моделирование пассажиропотока осуществляется при помощи регрессионных моделей, которые в качестве независимых переменных могут использовать довольно широкий спектр различных показателей: от численности агломерации до объема валового регионального продукта (ВРП) [16]. В зарубежной практике при моделировании пассажиропотоков используются также такие переменные, как размер туристического рынка агломерации, численность занятого населения, среднее и минимальное время поездки, средняя и минимальная стоимость поездки, количество автомобилей и пр. [12].

Однако, по мнению автора, визуальный анализ данных пассажиропотока является более перспективным, поскольку ежедневно собирается большой объем данных о перемещении пассажиров, который необходимо правильно и быстро интерпретировать.

В научной литературе существует большое разнообразие определений визуализации данных. Для начала определим, что понимается под визуальной аналитикой. Так, она представляет собой методы автоматического анализа с интерактивной визуализацией для эффективного понимания, обоснования и принятия решений на основе больших и сложных наборов данных [24].

В свою очередь визуализация данных представляет собой часть визуальной аналитики, которая использует компьютерные интерактивные визуальные представления данных для усиления познания с целью принятия решений или объяснения [6]. В зависимости от целей применяются различные подходы к визуализации данных [6, 13]:

- научная визуализация, которая представляет данные с заданной пространственной информации;
- визуализация абстрактных данных, которые не основаны на физических данных;
- географическая визуализация, которая представляет как абстрактные, так и пространственные данные.

Важно отметить, что визуализация данных осуществляется в несколько этапов: преобразование данных, визуальное отображение и преобразование вида. Соотношение этих этапов представлено на нижеследующем рисунке (рис. 1).



Рис. 1. Этапы визуализации данных. Источник: составлено автором на основе [6]

В рамках настоящего исследования наибольший интерес представляет именно визуализация пассажиропотоков. В целом, в визуальном анализе мобильности пассажиров большое внимание уделяется

анализу и мониторингу городских транспортных пассажиропотоков. На протяжении многих лет для этих целей использовались геоинформационные системы (ГИС) и традиционные методы визуализации (гистограммы, линейные графики, картографические карты в виде тепловых карт и пр.) [22]. Однако, уже в 2000 несколько ученых доказали, что для управления данными динамичных явлений, которыми и являются пассажиропотоки, ГИС имеет ряд ограничений. Впервые был предложен прототип ГИС для визуализации данных о городском движении, для чего использовались различные методы интерактивной визуализации трафика [7]:

- пространственная аналитика с помощью двумерных карт;
- тематическая аналитика при помощи диаграмм с областями;
- временная аналитика с помощью линейных графиков, которые представляют временные ряды;
- агрегирование с использованием гистограмм.

Позже, в 2002 году появился первый инструмент визуализации данных о транспортных потоках, который не был основан на ГИС. Он имел название CubeView и позволял отображать видео трафика, карты с интенсивностью движения и пиковые станции на установленную дату или время [20]. Далее инструменты визуализации данных о движении стали все больше совершенствоваться: появились 3D-модели, роз-диаграммы, матрицы тепловых карт, семантическое масштабирование, изоповерхность, техники кластеризации и многое другое.

Примеры использования названных инструментов визуализации данных движения представлены на нижеследующем рисунке (рис. 2).

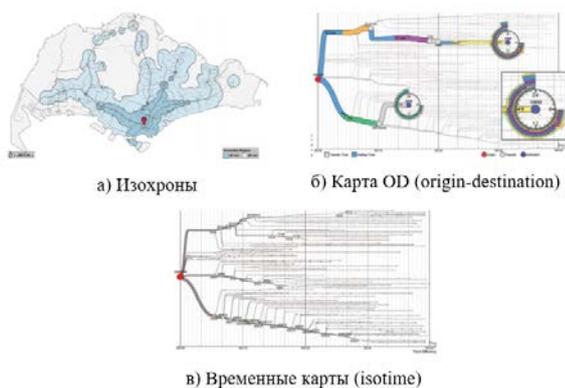


Рис. 2. Примеры визуализации данных. Источник: [26]

Однако, по мнению автора данные примеры использования методов визуального анализ данных являются неинформативными и сложными для восприятия. В следующем разделе будут рассмотрены более продвинутые инструменты визуализации данных движения и пассажиропотоков, в частности Kepler.gl.

Немного более подробно необходимо остановиться на методах интерактивной визуализации данных движения. В настоящее время существует множество таких методов. Первой группой методов является

универсальная мультивариативная визуализация, которая включает [11]:

- представление данных при помощи значков, отражающих одномерную информацию;
- семантические исследования, которые используют доступную контекстную информацию о движении объекта в дополнение к данным о его местоположении;
- облака слов, представленные в виде геопространственных тегов и ключевых слов.

Вторая группа методов включает визуализацию, ориентированную на низкоразмерные данные. В данном случае выделяются пространственно-временные кубы, штабелирование, тепловые карты, сетки и временные ряды. Также, к данной группе методов визуализации данных движения относятся и карты плотности, которые используются для обнаружения геопространственных участков пространства с критическими значениями плотности [19].

Также, существует и третья группа методов визуального анализа – это визуализация, ориентированная на данные большого размера. Чаще всего эти методы работают и обычно существуют в негеографическом пространстве, поскольку анализируются многомерные данные. Поэтому используются такие методы, как уменьшение размерности, построение проекции, иерархии, пиксельных карт, параллельных координат, а также радиальных графиков [11].

Можно выделить и четвертую группу методов визуального анализа геопространственных данных – это универсальная многомерная визуализация. Специалисты выделяют кластеризацию, графики разброса, визуализацию потоков на основе пространственных представлений и визуализацию потоков на основе непространственных представлений (например, метод «отправления-назначения»).

IV. ИНСТРУМЕНТЫ ВИЗУАЛЬНОГО АНАЛИЗА ДАННЫХ

В настоящее время для визуального анализа данных существует большое число различных инструментов, среди которых приложения Kepler.gl, Carto и Locale.ai. Так, Kepler.gl была разработана компанией Uber Engineering для аналитики большого объема геопространственных данных о дорожном движении, которые собирались автомобилями [4]. В свою очередь, Carto разработано компанией CARTODB Inc. и является платформой для предоставления инструментов ГИС, веб-картографии и анализа пространственных данных [3]. А приложение Locale.ai, разработанное индийской компанией LocateAI Technologies Private Limited, предоставляет возможность оперативной аналитики данных о местоположении [5].

Данные инструменты широко используются крупными компаниями в различных целях. Так, самым известным примером визуализации данных при помощи Kepler.gl является компания Uber Engineering, которая широко использовала этот инструмент для анализа дорожного движения в различных городах мира (к примеру, анализ опасных участков дорог в Манхэттане). Инструмент Carto используется большим числом

различных компаний, таких как JLL (услуги в сфере продажи коммерческой недвижимости), Mastercard (продукт Mastercard Retail Location Insights для оценки трендов в потреблении на уровне улиц), Vodafone (аналитика мобильных данных) и множество других компаний. А приложение Locale.ai использовалось мексиканской логистической компанией OTIF, лидирующими логистическими компаниями США и многими другими.

В нижеследующей таблице (табл. 1) представлен сравнительный анализ инструментов визуального анализа данных Kepler.gl, Carto и Locale.ai по таким параметрам, как целевая аудитория, направления использования и технические характеристики.

Таблица 1. Сравнительный анализ инструментов визуального анализа данных.

Инструмент	Целевая аудитория	Направления использования	Технические характеристики
Kepler.gl	Бизнес	Визуализация, анализ и сбор большого разнообразия данных: от карты землетрясений и численности населения до анализа поездок и интенсивности движения	Deck.gl и WebGL (API JavaScript), Mapbox GL
Carto	Бизнес и веб-разработчик и	Оптимизация логистики и цепочек поставок, поиск площадок, в том числе для розничной торговли, геомаркетинг, планирование и оптимизация трафика	CARTO Builder
Locale.ai	Бизнес	Геомаркетинг, анализ спроса и предложения, анализ трафика и поездок, анализ производительности статических объектов	Deck.gl, Nebula.gl и Mapbox GL

Источник: составлено автором на основе проведенного анализа.

Наравне с названными выше инструментами визуального анализа данных могут применяться и другие инструменты геопространственной аналитики. Например, ГИС QuantumGIS, TileMill или API Яндекс.Карт. Однако, функциональные возможности данных сервисов, по мнению автора, несколько ниже, чем у рассмотренных ранее инструментов.

Каждый из представленных инструментов геопространственной визуализации и аналитики имеет также свои минусы. Например, Kepler.gl работает с данными только в файлах форматов CSV, JSON и geoJSON. Это значительно усложняет аналитику большого объема данных, поскольку для каждого построения графика необходимо загружать файлы в названных форматах. А это значит, что любая обработка данных должна осуществляться вне приложения Kepler.gl.

Недостатки есть и у инструмента Carto. Несмотря на

то, что у него есть большое число возможностей для анализа данных, он может производиться только на одном наборе данных за один раз. Получается, что для того, чтобы визуализировать данные в Carto их необходимо сначала обработать на сторонних ресурсах вне системы, что усложняет процесс анализа больших объемов данных. Среди недостатков Locale.ai можно выделить небольшой набор методов аналитики, в сравнении с другими инструментами геопространственной аналитики.

Таким образом, одним из наиболее продвинутых инструментов визуализации данных пассажиропотока является 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. Далее будут рассмотрены преимущества данного приложения в сравнении с прочими инструментами визуального анализа данных.

V. ПРЕИМУЩЕСТВА KEPLER.GL В СРАВНЕНИИ С ДРУГИМИ ИНСТРУМЕНТАМИ ВИЗУАЛЬНОГО АНАЛИЗА ДАННЫХ

В сравнении с другими инструментами визуального анализа данных Kepler.gl имеет ряд значимых преимуществ. Во-первых, данное приложение разработано на основе библиотеки визуализации крупномасштабных наборов данных deck.gl на программном интерфейсе приложения WebGL. Также в основе приложения лежит react – упаковщик для Mapbox GL (react-mapbox-gl). Эти инструменты поставляются командой Uber Data Visualisation в фреймворк с открытым исходным кодом Vis.js. Вычисления используют программное ускорение, что дает возможность обработки большого количества данных за очень короткое время. Именно благодаря техническим возможностям, Kepler.gl является более мощным инструментом, в сравнении со стандартными картографическими сервисами.

Во-вторых, приложение Kepler.gl использует большое число различных форматов визуализации, в числе которых слои, палитры, фильтры, стили карт, а также дополнительные настройки. Для начала рассмотрим основные виды слоев, которые используются для визуального анализа данных. Их описание представлено в нижеследующей таблице (табл. 2).

Таблица 2. Слои для визуализации в приложении Kepler.gl.

Название слоя	Описание
Point	Данные накладываются на карту в виде точек
S2 Layer	Данные представляются в виде столбчатых диаграмм
Icon	Данные представляются в виде иконок для дифференциации объектов на карте
Line	Данные представлены в виде отрезков, соединяющих две точки

Cluster	Данные агрегируются на основе геопространственного радиуса
Polygon	Данные представляются в виде многоугольников для отображения участков на карте
Hexbin	Данные отображаются в соответствии с их распределением, например, максимальное и минимальное, среднее, медианное и др.
Grid	Данные представляются в виде тепловой карты в формате 2D и столбчатых диаграмм в формате 3D
НЗ	Данные отображаются в виде гексагонов
Heatmap	Интенсивность данных представляется в географических точках при помощи разных цветов
Arc	Отражаются дистанции между двух точек в формате 3D
Trip layer	Данные о поездках, которые можно изобразить и анимировать

Источник: составлено автором на основе анализа функциональных возможностей приложения Kepler.gl.

Следующим форматом визуализации в приложении Kepler.gl являются фильтры. Так, данное приложение может представлять временные данные, что возможно за счет наличия в функциональных возможностях гибкой пространственно-временной конфигурации. Также, в Kepler.gl есть функция географической фильтрации с названием «чистка», которая позволяет отобразить данные в пределах выбранного географического радиуса. Такая функция рассматриваемого предложения является уникальной среди прочих инструментов визуального анализа данных.

Наравне с названными выше техническими характеристиками Kepler.gl существует возможность дополнительных настроек. К ним относятся:

- настройка всплывающих окон для получения дополнительной информации об объекте на карте и его детального описания;
- использование сплит метода для сравнения двух карт с одинаковым набором данных, но с различными слоями;
- использование функции Time Playback, которая демонстрирует изменение визуализации во времени.

Именно за счет того, что рассматриваемое приложение имеет открытый исходный код и большое число технических возможностей, Kepler.gl является самым продвинутым и совершенным инструментом для решения проблемы возникновения очередей и скопления большого числа пассажиров на железнодорожных станциях.

VI. ВИЗУАЛИЗАЦИЯ ДАННЫХ ПАССАЖИРОПОТОКОВ ПРИ ПОМОЩИ KEPLER.GL

Важно отметить, что к наиболее важным данным, необходимым для визуального анализа данных пассажиропотоков при помощи Kepler.gl, относятся:

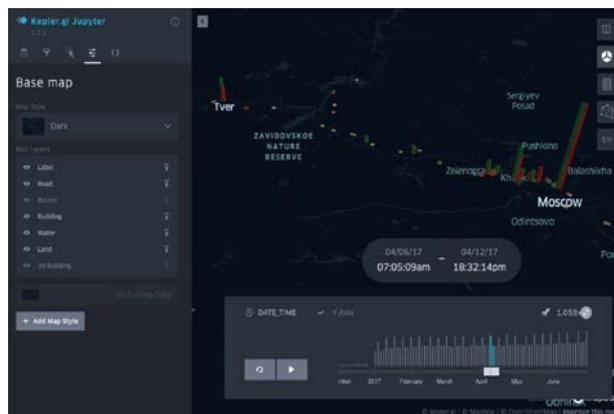
- дата и время прохода через турникет;
- номер турникета и его типа (на вход или на выход);
- билет по типу поезда: скорый или пассажирский;
- тип билета: льготный, взрослый, детский;

- тип билета: разовый «туда», абонемент обычный, разовый «туда-обратно», билет на число поездок, мастер-карта, АБОД, провожающий, БВД;
- носитель билета: карта или бумажный билет;
- станция покупки билета.

Также, при визуальном анализе пассажиропотока железнодорожных станций представленные выше данные могут сравниваться между собой. Так, можно сравнивать количество пассажиров в конкретную дату и время (вплоть до часов и минут), которые осуществляют вход на станцию или выход из нее. При этом, визуализация данных позволяет сравнивать не только данные о пассажиропотоке на одной станции, но и между различными станциями. Это позволяет определять наиболее загруженные пассажирами железнодорожные станции и, соответственно, принимать решения об управлении пассажиропотоком на них.

Для примера визуализации данных пассажиропотоков при помощи приложения Kepler.gl были использованы данные с турникетов на железнодорожных станциях по направлению движения «Москва-Ленинградская – Тверь» в период с 01.12.2016 г. по 30.06.2017 г. Всего за данный период зафиксировано 29 984 707 входов и выходов через турникеты, расположенные на 359 железнодорожных станциях.

Эти данные представлены на нижеследующем рисунке (рис. 3).



Примечание: красным цветом отмечены места прибытия, зеленым – отправления.

Рис. 3. Визуализация данных по маршруту «Москва-Ленинградская – Тверь» (столбчатый слой). Источник: составлено автором.

Для представленной визуализации данных были выбран столбчатый слой с данными о том, откуда и куда перемещались пассажиры, а также были настроены фильтры с датой и временем поездки. Также, данные можно представить при помощи другого слоя – arc (рис. 4).

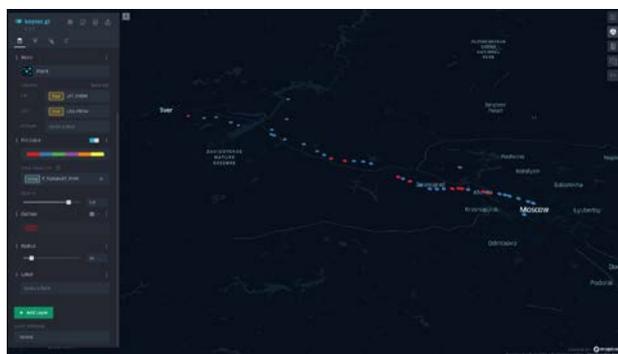


Примечание: красным цветом отмечены места прибытия, зеленым – отправления.

Рис. 4. Визуализация данных по маршруту «Москва-Ленинградская – Тверь» (arc слой). Источник: составлено автором.

На основе представленных выше визуальных представлений данных с турникетов на станциях о пассажиропотоке на направлении Москва-Ленинградская – Тверь» можно сделать два вывода. Во-первых, видно, что большая часть пассажиров перемещается между крупными населенными пунктами. Во-вторых, наблюдается превышение числа отправлений над числом прибытий.

Также, представленные данные о пассажиропотоке по маршруту «Москва-Ленинградская – Тверь» можно визуализировать при помощи других слоев. Они представлены на нижеследующем рисунке (рис. 5).



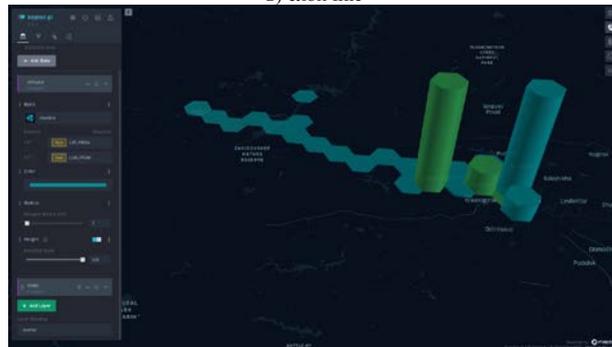
а) слой point



б) слой arc



в) слой line



г) слой hexbin



д) слой grid

Рис. 5. Визуализация данных по маршруту «Москва-Ленинградская – Тверь» при помощи различных слоев. Источник: составлено автором.

На рисунке число входов и выходов обозначается различными цветами. Так, число входов на рисунке 5а обозначено красным цветом, на рисунке 5б и 5в – бежевым цветом, на рисунке 5г – синим цветом, а на рисунке 5д – темно синими цветом. Получается, что визуализация данных пассажиропотока по маршруту «Москва-Ленинградская – Тверь» при помощи различных слоев инструмента Kepler.gl показывает, что по данному маршруту на железнодорожных станциях отмечается большее число входов, чем выходов. При этом, среди всех железнодорожных станций данного направления наиболее загруженной является станция в г. Москве «Ленинградский вокзал».

VII. ЗАКЛЮЧЕНИЕ

В настоящее время визуальный анализ данных является одним из наиболее перспективных методов оценки большого объема данных о пассажиропотоке железнодорожного транспорта. Актуальность такого метода с каждым годом все больше возрастает, что обусловлено широким распространением и

использованием цифровых технологий, таких как искусственный интеллект, большие данные и прочие.

Одним из наиболее продвинутых инструментов визуализации данных о пассажиропотоке железнодорожного транспорта для решения проблемы возникновения очередей и скопления большого числа пассажиров на железнодорожных станциях является приложение Kepler.gl. 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 по направлению «Москва-Ленинградская – Тверь» при помощи различных слоев. Использовались данные с турникетов на железнодорожных станциях по указанному маршруту о входах и выходах пассажиров. В результате приведенного анализа были сделаны три принципиально важных выводов. Во-первых, большая часть пассажиров перемещается между крупными населенными пунктами. Во-вторых, наблюдается превышение числа отправок над числом прибытий. В-третьих, среди всех станций маршрута движения пассажиров «Москва-Ленинградская – Тверь» наиболее загруженной является станция, расположенная в г. Москве на Ленинградском вокзале («Москва-Ленинградская»).

Качественные выводы, полученные в результате визуального анализа данных о пассажиропотоке по направлению «Москва-Ленинградская – Тверь» позволяют принимать управленческие решения по оптимизации пассажиропотока во избежание таких проблем, как возникновение больших очередей на железнодорожных станциях, их переполнение и создание заторов, большой поток пассажиров на направлениях пригородного сообщения, перегрузки железнодорожного транспорта во время часов пик, а также высокая загруженность определенных железнодорожных станций, в сравнении с другими станциями и другие проблемы.

БИБЛИОГРАФИЯ

- [1] Алешко Р. А. и др. Разработка методики визуализации и обработки геопространственных данных // Научная визуализация. – 2015. – Т. 7. – №. 1.
- [2] 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.
- [3] Официальный сайт Carto. URL: <https://carto.com/>
- [4] Официальный сайт Kepler.gl. URL: <https://kepler.gl/>
- [5] Официальный сайт Locale.ai. URL: <https://www.locale.ai/>
- [6] Card M. Readings in information visualization: using vision to think. – Morgan Kaufmann, 1999.
- [7] Claramunt, C.; Jiang, B.; Bargiela, A. A new framework for the integration, analysis and visualisation of urban traffic data within geographic information systems. Transp. Res. Part C Emerg. Technol. 2000, 167–184.
- [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] Gomes G. A. M. et al. Real-time discovery of hot routes on trajectory data streams using interactive visualization based on gpu // Computers & Graphics. – 2018. – Т. 76. – С. 129-141.
- [10] 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.
- [11] He J. et al. Diverse visualization techniques and methods of moving-object-trajectory data: A review // ISPRS International Journal of Geo-Information. – 2019. – Т. 8. – №. 2. – С. 63.
- [12] Hsu Y. T. et al. Forecasting High-speed Rail Ridership Using Aggregate Data: A case Revisit of High speed Rail in Taiwan // TRB 94th Annual Meeting Compendium of Papers. – 2015.
- [13] Kerren A. et al. (ed.). Information Visualization: Human-Centered Issues and Perspectives. – Springer, 2008. – Т. 4950.
- [14] Krause J. et al. Interactive visualization for real-time public transport journey planning // Proceedings of SIGRAD 2012; Interactive Visual Analysis of Data; November 29-30; 2012; Växjö; Sweden. – Linköping University Electronic Press, 2012. – №. 081. – С. 95-98.
- [15] Krüger R. Visual analytics of human mobility behavior: Dissertation. Institut für Visualisierung und Interaktive Systeme der Universität Stuttgart – 2017. 212 p.
- [16] Misharin A., Namiot D., Pokusaev O. On Processing of Correspondence Matrices in Transport Systems // 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). – IEEE, 2019. – С. 1-6.
- [17] Pensa S. Interactive Visualization Tool (InViTo) // Accessibility Instruments for Planning Practice. COST Office, (ISBN13: 978-989-20-3187-3 (hbk)). – 2012. – С. 978-989.
- [18] Rabah K. Convergence of AI, IoT, big data and blockchain: a review // The lake institute Journal. – 2018. – Т. 1. – №. 1. – С. 1-18.
- [19] Scheepens R. et al. Interactive visualization of multivariate trajectory data with density maps // 2011 IEEE Pacific Visualization Symposium. – IEEE, 2011. – С. 147-154.
- [20] Schreck T. et al. Visual cluster analysis of trajectory data with interactive kohonen maps // Information Visualization. – 2009. – Т. 8. – №. 1. – С. 14-29.
- [21] Shekhar, S.; Lu, C.T.; Liu, R.P.; Zhou, C. Cube view: A system for traffic data visualization. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Singapore, 6 September 2002.
- [22] Sobral T., Galvão T., Borges J. Visualization of urban mobility data from intelligent transportation systems // Sensors. – 2019. – Т. 19. – №. 2.
- [23] Tao S., Rohde D., Corcoran J. Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap // Journal of Transport Geography. – 2014. – Т. 41. – С. 21-36.
- [24] Thomas J. J. Illuminating the path: the research and development agenda for visual analytics. – IEEE Computer Society, 2005.
- [25] Zeng L. et al. A passenger flow control method for subway network based on network controllability // Discrete Dynamics in Nature and Society. – 2018.
- [26] Zeng W. et al. Visualizing mobility of public transportation system // IEEE transactions on visualization and computer graphics. – 2014. – Т. 20. – №. 12. – С. 1833-1842.
- [27] Соколов И. А. и др. Искусственный интеллект как стратегический инструмент экономического развития страны и совершенствования ее государственного

управления. Часть 1. Опыт Великобритании и США
//International Journal of Open Information
Technologies. – 2017. – Т. 5. – №. 9. – С. 57-75.

- [28] 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.
- [29] 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.

Visual analysis of railway passenger traffic data

Stepan Medvedenko, Dmitry Namiot

Abstract — This article discusses the features of visual analysis of railway passenger traffic data using the Kepler.gl visualization application. First of all, we consider the problems associated with the growth of passenger traffic in cities, which arise as a result of an increase in population, growth of cities, and the development of transport infrastructure. As a result, we substantiate the necessity of effective management of the passenger traffic of railway transport and the solution of the existing problems, including through the visual analysis of data. Next, we carry out a comparative description of existing visualization tools and it is proved that the most advanced of them is the Kepler.gl application. Due to its open-source code and a large number of technical capabilities, it is the most advanced and perfect tool for solving the problem of queues and congestion of large numbers of passengers at railway stations. Using the Kepler.gl application, we visualize the data collected from the passageways through the turnstiles at the railway stations. Based on the visualization of these data, it was possible to come to the conclusion that the main passenger flows move between large cities, and the number of departures from them exceeds the number of arrivals.

Key words — data visualization, visual data analysis, visualization of passenger traffic, railway transport

REFERENCES

- [1] Aleshko R. A. i dr. Razrabotka metodiki vizualizacii i obrabotki geoprostranstvennyh dannyh //Nauchnaja vizualizacija. – 2015. – T. 7. – #. 1.
- [2] Beljakov S. L., Beljakova M. L., Savel'eva M. N. Adaptivnaja k izmeneniju struktury bazy dannyh vizualizacija prostranstvennyh dannyh //Pribory i sistemy. Upravlenie, kontrol', diagnostika. – 2016. – #. 1. – S. 25-32.
- [3] Oficial'nyj sajt Carto. URL: <https://carto.com/>
- [4] Oficial'nyj sajt Kepler.gl. URL: <https://kepler.gl/>
- [5] Oficial'nyj sajt Locale.ai. URL: <https://www.locale.ai/>
- [6] Card M. Readings in information visualization: using vision to think. – Morgan Kaufmann, 1999.
- [7] Claramunt, C.; Jiang, B.; Bargiela, A. A new framework for the integration, analysis and visualisation of urban traffic data within geographic information systems. *Transp. Res. Part C Emerg. Technol.* 2000, 167–184.
- [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. – S. 290-291.
- [9] Gomes G. A. M. et al. Real-time discovery of hot routes on trajectory data streams using interactive visualization based on gpu //Computers & Graphics. – 2018. – T. 76. – S. 129-141.
- [10] 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. – S. 138-166.
- [11] He J. et al. Diverse visualization techniques and methods of moving-object-trajectory data: A review //ISPRS International Journal of Geo-Information. – 2019. – T. 8. – #. 2. – S. 63.
- [12] Hsu Y. T. et al. Forecasting High-speed Rail Ridership Using Aggregate Data: A case Revisit of High speed Rail in Taiwan //TRB 94th Annual Meeting Compendium of Papers. – 2015.
- [13] Kerren A. et al. (ed.). Information Visualization: Human-Centered Issues and Perspectives. – Springer, 2008. – T. 4950.
- [14] Krause J. et al. Interactive visualization for real-time public transport journey planning //Proceedings of SIGRAD 2012; Interactive Visual Analysis of Data; November 29-30; 2012; Växjö; Sweden. – Linköping University Electronic Press, 2012. – #. 081. – S. 95-98.
- [15] Krüger R. Visual analytics of human mobility behavior: Dissertation. Institut für Visualisierung und Interaktive Systeme der Universität Stuttgart – 2017. 212 p.
- [16] Misharin A., Namiot D., Pokusaev O. On Processing of Correspondence Matrices in Transport Systems //2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). – IEEE, 2019. – S. 1-6.
- [17] Pensa S. Interactive Visualization Tool (InViTo) //Accessibility Instruments for Planning Practice. COST Office, (ISBN13: 978-989-20-3187-3 (hbk). – 2012. – S. 978-989.
- [18] Rabah K. Convergence of AI, IoT, big data and blockchain: a review //The lake institute Journal. – 2018. – T. 1. – #. 1. – S. 1-18.
- [19] Scheepens R. et al. Interactive visualization of multivariate trajectory data with density maps //2011 IEEE pacific visualization symposium. – IEEE, 2011. – S. 147-154.
- [20] Schreck T. et al. Visual cluster analysis of trajectory data with interactive kohonen maps //Information Visualization. – 2009. – T. 8. – #. 1. – S. 14-29.
- [21] Shekhar, S.; Lu, C.T.; Liu, R.P.; Zhou, C. Cube view: A system for traffic data visualization. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Singapore, 6 September 2002.
- [22] Sobral T., Galvão T., Borges J. Visualization of urban mobility data from intelligent transportation systems //Sensors. – 2019. – T. 19. – #. 2.
- [23] Tao S., Rohde D., Corcoran J. Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap //Journal of Transport Geography. – 2014. – T. 41. – S. 21-36.
- [24] Thomas J. J. Illuminating the path: the research and development agenda for visual analytics. – IEEE Computer Society, 2005.
- [25] Zeng L. et al. A passenger flow control method for subway network based on network controllability //Discrete Dynamics in Nature and Society. – 2018.
- [26] Zeng W. et al. Visualizing mobility of public transportation system //IEEE transactions on visualization and computer graphics. – 2014. – T. 20. – #. 12. – S. 1833-1842.
- [27] Sokolov I. A. i dr. Iskusstvennyj intellekt kak strategicheskij instrument jekonomicheskogo razvitija strany i

sovershenstvovaniya ee gosudarstvennogo upravlenija. Chast' 1. Opyt Velikobritanii i SShA //International Journal of Open Information Technologies. – 2017. – T. 5. – #. 9. – S. 57-75.

[28] Kuprijanovskij V. P. i dr. Pravitel'stvo, promyshlennost', logistika, innovacii i intellektual'naja mobil'nost' v cifrovoj jekonomike //Sovremennye informacionnye tehnologii i IT-obrazovanie. – 2017. – T. 13. – #. 1. – S. 74-96.

[29] 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. – S. 012012.