

МЕТОДИКИ ОЦЕНКИ ПАССАЖИРОПОТОКА НА КОММЕРЧЕСКИХ МАРШРУТАХ САНКТ-ПЕТЕРБУРГА

Филатова А.А., кафедра информационно-аналитических систем
СПбГУ, anastasia.a.filatova@gmail.com

Аннотация

В данной работе описываются две методики оценки пассажиропотока на коммерческом транспорте Санкт-Петербурга, алгоритмы, которые лежат в их основе, и детали их реализации. Первый алгоритм базируется на идее использования информации о пассажиропотоке на муниципальном транспорте, который следует по похожим маршрутам. В основе второго алгоритма лежит обработка данных о транзакциях оплаты проезда на коммерческих маршрутах с использованием смарт-карт «Подорожник». Для каждой из методик приводится анализ их применимости.

Введение

Характерной особенностью Санкт-Петербурга является широко развитая сеть коммерческого транспорта. На сегодняшний день все больше людей предпочитают коммерческие маршруты муниципальным, потому что они быстры и мобильны, а для пассажиров, не имеющих социальных льгот, стоимость проезда в них сравнима со стоимостью проезда на других видах транспорта. В связи с этим возникает необходимость анализа существующей системы коммерческих маршрутов для улучшения ее качества, а в частности, оценки пассажиропотока и построения алгоритмов его моделирования, чему и посвящена моя работа.

В Санкт-Петербурге для городских маршрутов был разработан ряд систем, позволяющих упростить как пользование транспортом, так и анализ существующей транспортной системы. В первую очередь это введение возможности оплаты проезда с использованием бесконтактных смарт-карт (БСК). Примерами бесконтактных смарт-карт являются: электронная карта-кошелек «Подорожник», студенческие и пенсионные карты. Информация о транзакциях оплаты проезда смарт-картами учитывается системой электронного контроля оплаты проезда (СЭКОП) и может быть использована для дальнейшего анализа. Также для удобства отслеживания транспортных средств на муниципальных маршрутах (т.е. на маршрутах, на которых действует система льготного проезда для

определенных социальных групп пассажиров) была введена автоматизированная система управления городским пассажирским транспортом (АСУГПТ). Она позволяет определять местоположение конкретного транспортного средства на маршруте в определенный момент времени. Это происходит благодаря специальным датчикам, установленным на борту.

Существуют важные особенности, которые не позволяют провести анализ характеристик коммерческих маршрутов так же, как это сделано для муниципального транспорта. Например, на коммерческих маршрутах допустима оплата проезда смарт-картами только в виде электронного кошелька «Подорожник», да и в целом, процент людей, оплачивающих проезд смарт-картами на них, значительно ниже, чем на муниципальных. Еще одной особенностью является невозможность установить на коммерческих маршрутах датчики АСУГПТ и, как следствие, нет возможности отслеживать коммерческие транспортные средства на маршрутах. В связи с этим возникает необходимость разработки других методов анализа характеристик коммерческих маршрутов и, в частности, пассажиропотока.

Описание используемых данных

Для решения задачи оценки пассажиропотока на коммерческом транспорте СПб ГКУ «Организатор перевозок» предоставил следующие данные. Они были собраны в период август-октябрь 2018 года.

- Данные транзакций оплаты проезда смарт-картами «Подорожник» на коммерческих маршрутах. Содержат более 1.2 млн. записей, получены из системы электронного контроля оплаты проезда.
- Данные о порядке прохождения остановок муниципальными маршрутами. Содержат более 2 млн. записей, получены из автоматизированной системы управления городским пассажирским транспортом.
- Данные наблюдателей. Содержат информацию о количестве пассажиров на коммерческих транспортных средствах в определенные дни на 30 маршрутах.
- Список всех городских остановок. Содержит информацию об идентификаторах остановок, их названиях и координатах, а также кластерах, которым они принадлежат. Здесь и далее кластерами будем называть группы остановок, которые расположены в непосредственной близости друг к другу.

Так как большинство полученных данных – это показания датчиков, и, следовательно, содержат пропуски и ошибки, была проведена базовая предобработка таких данных. Кроме того, записи были обработаны и структурированы для удобства их дальнейшего использования в алгоритмах.

Ввиду того, что не всю информацию о коммерческих маршрутах можно было получить из автоматизированных систем сбора, потребовалось дополнительно вручную собрать данные об остановках на исследуемых коммерческих маршрутах, расстоянии между ними, кластерах, которым они принадлежат, и времени прохождения маршрутов коммерческими транспортными средствами. Собранные данные также были структурированы.

Разведочный анализ данных

Для того чтобы узнать структуру данных и выявить базовые зависимости, был проведен разведочный анализ данных. Он показал, что распределения по часам среднего количества пассажиров, которые оплачивают проезд на маршрутках смарт-картами, в будние и выходные дни в точности совпадают с аналогичными распределениями на городском транспорте (рис. 1, 2).



Рисунок 1. Распределение по часам среднего количества транзакций оплаты проезда смарт-картами «Подорожник» в будние дни на коммерческих и муниципальных маршрутах



Рисунок 2. Распределение по часам среднего количества транзакций оплаты проезда смарт-картами «Подорожник» в выходные дни на коммерческих и муниципальных маршрутах

Для коммерческих и муниципальных маршрутов характерны одинаковые пики в будние дни в часы, когда пассажиры, работающие по пятидневной системе, едут на работу и с работы и равномерно распределенное число транзакций на выходных с 9 утра до 7 вечера. Это наблюдение легло в основу одной из методик.

Методика на основе анализа характеристик похожих муниципальных маршрутов

В основе первой методики лежит предположение, которое базируется на результатах разведочного анализа данных. Так как распределения количества пассажиров на коммерческих и муниципальных маршрутах совпадают, будем считать, что коммерческие и муниципальные транспортные средства, следующие по одному маршруту в одно и то же время, загружены одинаково.

Тогда для оценки пассажиропотока можно сделать следующее:

1. Найти муниципальные маршруты, которые похожи на исследуемый коммерческий маршрут.
2. Для таких муниципальных маршрутов вычислить коэффициент загрузки – процент, показывающий отношение реального количества пассажиров к максимально возможному.
3. Использовать полученное значение коэффициента загрузки для расчета пассажиропотока на исследуемом коммерческом маршруте.

Для реализации данной методики необходимо было разработать и реализовать алгоритм поиска "похожих" маршрутов.

Будем называть маршруты тем более похожими, чем больше у них общих перегонов. Общие перегоны – это перегоны, начальные остановки которых принадлежат одному кластеру и конечные остановки которых принадлежат одному кластеру (**рис. 3**).

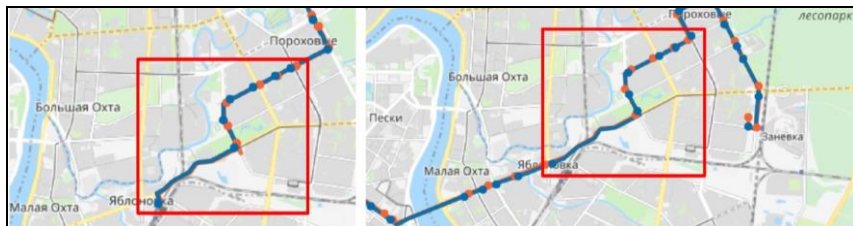


Рисунок 3. Пример похожих маршрутов. Слева представлен коммерческий маршрут №92, а справа – часть автобусного маршрута №27.
Цепочка общих перегонов выделена красным

Задача поиска наиболее похожих маршрутов свелась к задаче поиска непересекающихся общих цепочек перегонов у двух маршрутов. В контексте этой задачи наиболее похожими будут такие маршруты, у которых суммарная длина получившихся цепочек максимальна.

Итоговый алгоритм поиска похожих маршрутов на вход принимает информацию об исследуемом коммерческом маршруте и параметр, показывающий, какое количество наиболее похожих муниципальных маршрутов следует вывести. Далее происходит поочередное сравнение коммерческого и муниципальных маршрутов и формируется список, отсортированный по убыванию суммарных длин цепочек общих перегонов каждой пары таких маршрутов. В качестве результата алгоритм выводит интересующее количество наиболее похожих маршрутов и указывает процент «похожести» (т.е. отношение количества общих перегонов к числу перегонов на исследуемом коммерческом маршруте). Результат представляется в виде небольшого текстового отчета (**рис. 4**).

Для К-68 найдены следующие наиболее похожие:

Прямое направление К-68 совпадает с обратным направлением автобуса №68, процент совпадения: 100%

Прямое направление К-68 совпадает с прямым направлением троллейбуса №37, процент совпадения: 83%

Рисунок 4. Результат работы алгоритма поиска похожих маршрутов.
Выведены два наиболее похожих муниципальных маршрута для К-68.

Методика на основе данных о транзакциях оплаты проезда смарт-картами на коммерческих маршрутах

Вторая методика основана на анализе транзакций оплаты проезда на маршрутах смарт-картами Подорожник. Так как есть возможность автоматизированного сбора таких данных о транзакциях, возникла идея использовать их для оценки пассажиропотока. Но для коммерческих маршрутов отсутствует информация о месте проведения транзакций, поэтому необходимо было сначала научиться ее восстанавливать. Кроме того, количество людей на коммерческих маршрутах, которые оплачивают проезд смарт-картами, значительно меньше реального количества пассажиров. Это тоже необходимо было учесть. Поэтому разработанная методика включает в себя следующие шаги:

1. Восстановление информации о месте посадки пассажиров, плативших БСК. Фактически, это восстановление информации о месте проведения транзакции.
2. Восстановление информации о месте выхода пассажиров, плативших БСК.
3. Построение матриц отправления-назначения для подсчёта пассажиропотока среди таких пассажиров.
4. Оценка реального числа пассажиров с использованием коэффициентов различия между количеством пассажиров, оплатившими проезд смарт-картами, и реальным количеством пассажиров. Они были рассчитаны по данным наблюдателей.

Так как рассчитать места выхода пассажиров можно с использованием существующего метода корреспонденций [1], основную сложность представляла разработка и реализация алгоритма восстановления мест посадки пассажиров.

Разработанный алгоритм восстановления мест посадки пассажиров основывается на идеях, описанных в работе [2] и адаптирован под транспортную систему Санкт-Петербурга. Он состоит из четырех шагов:

1. Разделение транзакций на различные поездки. Каждой такой поездке присваивается направление (прямое или обратное).
2. Кластеризация транзакционных записей. К одному кластеру будут отнесены последовательные записи, между которыми прошло менее 20 секунд. Каждый такой кластер представляет группу пассажиров, которые сели в одном месте.

3. Сопоставление кластеров и остановок на маршруте.
4. Корректировка полученных данных. Необходимость её обусловлена тем, что водители коммерческих транспортных средств начинают набирать новых пассажиров сразу после высадки старых на конечной остановке. В таких случаях разделение поездок может не произойти на первом шаге.

Результатом работы алгоритма стало восстановление для каждой транзакции идентификатора остановки, на которой садился пассажир, ее совершивший.

Анализ применимости методик

Первая методика основывается на проверенных данных о пассажиропотоке на муниципальных маршрутах и алгоритме поиска «похожих» маршрутов, который был хорошо протестирован, поэтому она достаточно надежна и проста в применении.

Методика на основе транзакций оплаты проезда смарт-картами требует качественно подготовленных и очищенных транзакционных данных, поэтому при ее использовании следует уделять особое внимание обработке входных данных. Кроме того, данная методика использует информацию наблюдателей, сбор которой сопряжен с высокими накладными расходами.

Заключение

В работе описаны две методики оценки пассажиропотока на коммерческих маршрутах и анализ их применимости. Созданные датасеты с данными о муниципальных и коммерческих маршрутах выложены в открытый доступ. Ознакомиться с ними и скачать их можно по ссылке: https://yadi.sk/d/u_9johMiW7FFQA. Реализация алгоритмов с подробным описанием и комментариями доступна в github-репозитории по ссылке: <https://github.com/nast1415/vkr2019>.

Литература

1. Natalia Grafееva, Elena Mikhailova, Elena Nogova, Innokenty Tretyakov, Passenger Traffic Analysys Based on St. Petersburg Public Transport, 17th International Multidisciplinary Scientific GeoConference: Informatics, Geoinformatics and Remote Sensing, Issue 21, SGEM 2017; Albena; Bulgaria; Volume 17, 2017, Pages 509-516
2. Chen Z.; Fan W. «Extracting bus transit boarding stop information using smart card transaction data», 2018