



МАРШРУТИЗАЦИЯ ТРАНСПОРТА ПРИ НАЛИЧИИ ОПАСНЫХ УЧАСТКОВ НА ДОРОГЕ (на примере города Спрингфилд, Массачусетс)

Герштейн А. М.¹, аспирант, ✉ ArkadyGer@gmail.com

Терехов А. Н.¹, доктор физико-математических наук, профессор, a.terekhov@spbu.ru

¹ Санкт-Петербургский государственный университет,
Университетская наб., д. 7–9, 199034, Санкт-Петербург, Россия

Аннотация

В работе [1] найдены статистически значимые кластеры дорожно-транспортных происшествий (ДТП), которые можно интерпретировать как участки повышенной опасности ДТП (УПО).

В данной работе, продолжающей исследования, проведенные в [1], рассмотрен простой способ обхода УПО при маршрутизации транспорта по критерию общей длины пути, заключающийся в том, что атрибуту длины каждой грани дорожного графа, ведущей к УПО, присваивается очень большое число, что делает эту грань практически непроходимой для алгоритма маршрутизации. Численные расчеты, для которых используется дорожная карта Спрингфилда (Массачусетс) и данные об УПО в Спрингфилде, показывают, что для маршрутов, чьи начальные и конечные пункты совпадают, маршрутизация с обходом УПО увеличивает протяженность маршрута относительно первоначальной¹, вычисленной без учета УПО. В пределе среднее отношение длин обоих маршрутов стремится для Спрингфилда к 1,04.

Для проверки эффективности маршрутизации введен новый показатель — относительный риск ДТП, равный отношению числа ДТП вдоль маршрута, учитывающего УПО, к числу ДТП, подсчитанных вдоль исходного, построенного без учета УПО маршрута.

Показано, что при использовании алгоритма обхода кластеров ДТП для маршрутов длиной более 4 км. средний относительный риск ДТП снижается на величину порядка 16 % при увеличении длины маршрута в среднем на 8%.

Ключевые слова: маршрутизация транспорта, относительный риск ДТП, алгоритм Дейкстры, транспорт, кластер ДТП, DBSCAN, статистическое испытание Монте-Карло, Массачусетс.

Цитирование: Герштейн А. М., Терехов А. Н. Маршрутизация транспорта при наличии опасных участков на дороге (на примере города Спрингфилд, Массачусетс) // Компьютерные инструменты в образовании. 2022. № 2. С. 5–18. doi: 10.32603/2071-2340-2022-2-5-18

¹ Причем, чем меньше первоначальная протяженность, тем больше увеличивается длина маршрута, построенного с учетом УПО.

1. ВВЕДЕНИЕ

Задача построения оптимального в том или ином смысле маршрута для индивидуального транспорта относится к числу хорошо изученных и давно решенных. Если дорожная сеть представлена двунаправленным мультиграфом (то есть, графом, где через две любые вершины могут проходить несколько граней, а каждая грань может быть одно- или двунаправленной), то маршрутизация проводится с помощью алгоритма Дейкстры [7] или Беллмана-Форда [8]. Каждой грани достаточно сопоставить число (для алгоритма Дейкстры — обязательно неотрицательное), и алгоритмы построят оптимальный (в смысле минимизации суммы таких чисел) маршрут.

Для обычных задач маршрутизации грани приписывается либо ее длина, либо время задержки — гораздо более сложный параметр, зависящий не только от геометрии дорожной сети, но и от текущего состояния трассы: наличия дорожных пробок, погодных условий и пр.

К сожалению, в обычной маршрутизации весьма редко учитываются другие важные факторы, в частности участки дороги с компактными участками повышенной опасности УПО (см. [1]) или с большим количеством серьезных (то есть приведших к травмам или летальным исходам) ДТП, находящихся в пределах одного дорожного сегмента (грани). Есть работа [2], где сделана попытка создания интегрированного признака, приписываемого каждой грани дорожного графа и содержащего некую взвешенную сумму времени задержки и показателя опасности ДТП для каждой грани. Такой подход весьма трудоемок и требует отдельных и весьма непростых расчетов для каждой грани графа, представляющего дорожную сеть. Между тем, существует простой прием, позволяющий обойти опасные участки, помечая их как «непроходимые».

Представим себе граф дорожной сети, в котором каждой грани приписана в качестве атрибута, используемого алгоритмом маршрутизации, ее длина. Тогда граням, ведущим к участку повышенной опасности (УПО), можно приписать бесконечную или, что более практично, очень большую (на несколько порядков больше длины самой протяженной грани) длину, и тогда алгоритм маршрутизации автоматически обойдет опасный участок дороги.

2. ИНСТРУМЕНТЫ И ДАННЫЕ

Применим этот простой прием для дорожной сети города Спрингфилд (Springfield, MA), извлеченной из глобальной карты земной поверхности OpenStreetMap (OSM)². Карты OSM, полученные в результате многолетней работы некоммерческого вики-проекта, могут быть разного качества в зависимости от местности³, но для целей исследования алгоритмов маршрутизации (а как показано в [3] на примере дорожной сети Вены, также и для реальной маршрутизации) они вполне подходят.

Единственная трудность состоит в извлечении дорожной сети для заданного участка поверхности. Когда для этого использовался OSM API, задача была далека от тривиальной. К счастью, недавно появившаяся библиотека OSMnx для языка Питон (см. подробности в [4]) упростила эту задачу до пары-другой строчек кода. Так, для получения карты любого

² См. <https://www.openstreetmap.org/>

³ Существуют многочисленные средства проверки актуальности и точности OSM-карт (см. https://wiki.openstreetmap.org/wiki/Quality_assurance). Так, например, сайт <https://is-osm-uptodate.frafra.eu/> позволяет посмотреть последние изменения карт и их прошлые редакции. Карта Спрингфилда согласно этому сайту создана в 2014 году, последние изменения датируются 2021 годом, что говорит об ее актуальности.

города достаточно указать сам город, штат (для США) и страну. Например, дорожная сеть Спрингфилда загружается с помощью функции `graph_from_place()` пакета `OSMnx`:

```
import osmnx as ox
ox.config(use_cache=True, log_console=True, all_oneway=False)
G = ox.graph_from_place('SPRINGFIELD, MA, USA', network_type='drive')
```

Остается только сохранить объект `G`, содержащий дорожную сеть, на диске для дальнейшего использования.

Объект `G` хранит сеть в виде МультиДиГрафа (`MultiDiGraph`), то есть графа с возможностью существования многих одно- и двунаправленных граней, проходящих через одну и ту же пару вершин. Такой формат графа не всегда удобен для геоинформационных систем (ГИС), подобных `QGIS` или `ArcGIS`, поэтому в пакете `OSMnx` существует возможность сохранения в “.shp” файлах, стандартных для ГИС, отдельно граней и вершин МультиДиГрафа, которой мы будем широко пользоваться для определения граней, ведущих к УПО.

Нам также понадобятся данные о ДТП в штате Массачусетс за 2013–2018 годы, которые можно получить в текстовом формате “.csv” на сайте портала `MassDOT`⁴.



Рис. 1. Сетка из равноотстоящих точек, наложенная на карту Спрингфилда

3. МЕТОД

Для исследования эффективности маршрутизации при наличии опасных участков дорожной сети, зададим сетку равноотстоящих точек и наложим ее на карту Спрингфилда (рис. 1). На карте поместилось 95 узлов сетки с расстоянием около 830 м между соседними

⁴ <https://massdot-impact-crashes-vhb.opendata.arcgis.com/search>

узлами, что дает $95 \times 94 = 8930$ различных направленных маршрутов⁵. Наши дальнейшие действия можно уложить в следующую схему:

1. Объединим все случившиеся в Спрингфилде в 2013–2018 годах серьезные ДТП и дорожный граф, извлеченный из OSM. Заметим, что координаты ДТП и дорожный граф получены из независимых источников, поэтому точки ДТП могут не совпадать с гранями графа. Следовательно, нужна процедура, которая проецирует каждое ДТП на ближайшую грань. В результате этой операции каждой грани графа будет сопоставлено неотрицательное число ДТП, случившихся на ее протяжении.

2. Подготовим две версии дорожного графа: G0 и G1. Версия G0 будет содержать исходный граф, извлеченный из OSM. В версии G1 атрибут, в котором хранится длина грани, будет для некоторых граней заменен очень большим числом, что делает такие грани практически непроходимыми (при наличии альтернатив) для алгоритма маршрутизации.

3. Пользуясь сеткой, наложенной на карту Спрингфилда, построим все возможные маршруты для графов G0 и G1. Далее выберем из всех маршрутов, соединяющих одни и те же точки в том же направлении, только отличающиеся для G0 и G1, то есть те, которые затронуты изменениями в графе G1.

4. Для всех выбранных направленных маршрутов сравним их длину и число вершин (по данным, хранящимся в графе G0), то есть, посмотрим, как меняется длина и «дробность» маршрута при обходе препятствий.

5. Для всех выбранных направленных маршрутов сравним числа ДТП вдоль маршрута, то есть посмотрим, как меняется относительный риск ДТП при обходе препятствий. Под относительным риском здесь понимается отношение $\text{ДТП}_1/\text{ДТП}_0$, где ДТП_0 — число ДТП, подсчитанное вдоль оригинального маршрута на графе G0, а ДТП_1 — число ДТП вдоль маршрута, соединяющего те же начальную и конечную точки, но на графе G1.

4. ОБХОД УЧАСТКОВ ПОВЫШЕННОЙ ОПАСНОСТИ (УПО)

В этом разделе мы попробуем строить маршруты транспортных средств так, чтобы обойти участки повышенной опасности (УПО) или статистически значимые метакластеры ДТП, то есть кластеры, полученные в результате вторичной кластеризации кластеров ДТП, случившихся в Спрингфилде в 2013–2018 годах. Из всех таких вторичных кластеров были отобраны только те, что содержат кластеры ДТП не менее чем за три различных года (всего 31). См. подробности в [1].

Наша задача достаточно сложна, поэтому разумно будет разбить ее на несколько этапов, сохраняя (там, где это необходимо) промежуточные данные в файлах формата “.csv”:

1. Загрузить с диска два оригинальных дорожных графа и назвать их G0 и G1.

2. Найти все грани, ведущие к УПО (метакластерам). Эту задачу приходится выполнять вручную, поскольку кластер может быть по-разному расположен. Если кластер ДТП расположен на однонаправленной грани, достаточно просто присвоить атрибуту, хранящему длину грани, очень большое число и грань станет практически непроходимой для алгоритмов маршрутизации, то есть кластер (УПО) будет исключен из прокладываемого по графу G1 маршрута. Если же кластер находится на двунаправленной грани, задача модификации графа усложняется, поскольку нужно будет приписать кластер одному из

⁵ Здесь учитывается, что некоторые грани дорожного графа могут быть однонаправленными, так что маршрут $n \rightarrow m$ не идентичен маршруту $m \rightarrow n$.

направлений грани. Наконец, в случае кластера, расположенного на перекрестке, приходится «закрывать» все пути, ведущие к кластеру, потому что нет какой-то выделенной грани и ДТП происходят по всем направлениям. Таким образом, граф G_1 легко можно изменить, если заранее подготовить список «запретных» граней, рассматривая отдельно каждый метакластер и все грани, ведущие к нему.

3. Для всех граней графа G_1 , определенных в предыдущем пункте, заменим значение атрибута длины грани очень большим числом 10000000000, на много порядков большим, чем длина самой длинной грани в G_1 .

4. Для всех начальных и конечных пунктов, задаваемых сеткой из равноотстоящих точек (см. рис. 1), вычислим оптимальные маршруты и выделим из них те, которые различаются для графов G_0 и G_1 при одинаковых начальной и конечной точках маршрутов. В результате получаются 4822 таких маршрута, причем длина некоторых маршрутов (всего их 91, и это составляет менее 2 % от всех полученных) превышает 10000000000, что говорит о том, что в этих случаях алгоритм маршрутизации не нашел путей обхода препятствий. Исключив такие маршруты из наших данных, получим окончательно 4731 пары маршрутов на графах G_0 и G_1 , имеющих одинаковые начало и конец маршрута.

Эти оставшиеся маршруты могут быть отображены в виде точек на графике, где по оси X отложены первоначальные длины маршрутов без учета препятствий, а по оси Y — отношения длин маршрутов, вычисленных по графам G_1 и G_0 (рис. 2), и отношения числа вершин заданного маршрута, вычисленных по графу G_1 , к числу вершин, вычисленных по графу G_0 (рис. 3).

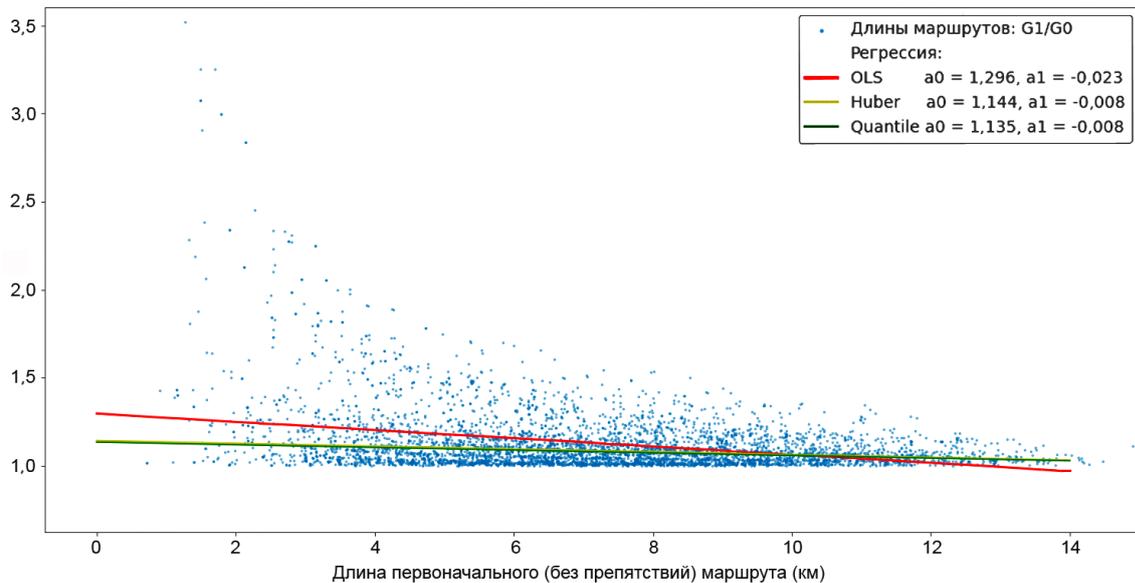


Рис. 2. Отношения длин маршрутов, вычисленных по графам G_1 (с учетом препятствий) и G_0 (без учета препятствий)

Чтобы лучше представлять, где сосредоточены основные данные, на рисунках 2 и 3 показаны различные регрессионные зависимости вида $y = a_0 + a_1 \cdot x$ отношений длин (рис. 2) или отношений числа вершин (рис. 3) от длины первоначального (без учета УПО) маршрута. Поскольку данные отличаются большим разбросом, на рисунках показаны также регрессионные прямые, полученные с помощью двух робастных (устойчивых к выборо-

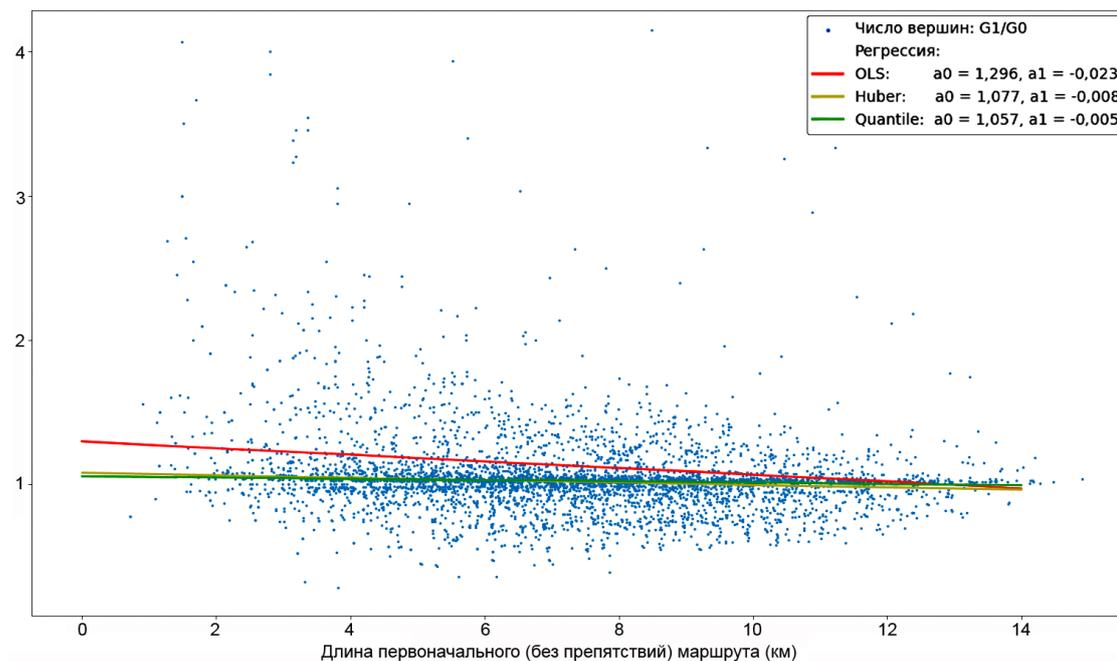


Рис. 3. Отношения числа вершин маршрутов, вычисленных по графам G1 (с учетом препятствий) и G0 (без учета препятствий)

сам) алгоритмов: регрессии Хубера⁶ (желтая прямая) и квантильной регрессии⁷ (зеленая прямая), позволяющей предсказывать произвольные квантили зависимой переменной. В нашем случае мы использовали предсказание для медианы (50 % квантиль). Коэффициенты регрессии показаны в пояснениях к рисункам.

Из диаграмм, показанных на рис. 2 и 3, хорошо видно, что длина пути в случае обхода препятствий всегда больше «оригинального» пути, построенного без учета препятствий, также хорошо видно, что отношение длины маршрутов G1/G0 уменьшается по мере увеличения длины оригинального пути.

Что же касается отношения числа вершин G1/G0, то для отдельных маршрутов оно может быть как больше, так и меньше единицы — в зависимости от того, на какую трассу «лег» обходной маршрут. Как показывают данные таблицы 2 (последний столбец), среднее отношение числа вершин стремится к единице для протяженных (более 6 км маршрутов)⁸ и увеличивается (видимо, за счет того, что обходные маршруты оказываются тут существенно длиннее) до 1,52 на самых коротких маршрутах.

Поскольку количество вершин оказывается малоинформативным и стремится к единице для достаточно протяженных маршрутов, основной интерес представляет проигрыш в расстоянии, возникающий при выборе альтернативного (с учетом обхода УПО) маршрута.

Приведем, прежде всего, общую описательную статистику по всем различающимся (при одних и тех же начальных и конечных координатах) маршрутам (таблица 1).

⁶ См. [9] и [10].

⁷ См. [10].

⁸ Отличие от единицы среднего отношения вершин G1/G0 для маршрутов с оригинальной (без учета УПО) протяженностью больше 13 км. связано, видимо, с небольшим размером выборки и вытекающей из этого потерей точности вычисления среднего.

Поскольку распределение отношений различных величин в нашем случае весьма далеки от нормального⁹, для вычисления доверительных интервалов всюду используется бутстреп-метод (см., например [5]). Конкретные расчеты выполнялись с помощью Python-пакета bootstrapped (<https://github.com/facebookarchive/bootstrapped>).

Таблица 1. Общая описательная статистика отношений длин маршрутов (Длина при обходе УПО/Первоначальная длина)

Показатель	Значение	95 % Доверительный интервал
Среднее	1,128	(1,122 – 1,133)
Медиана	1,07	(1,069 – 1,074)
Стандартное отклонение	0,18	(0,161 – 0,203)
Минимум	1,00	
Максимум	3,52	

Результаты расчетов для различной длины исходного (проложенного без учета УПО) маршрута показаны в таблице 2 и на рис. 4.

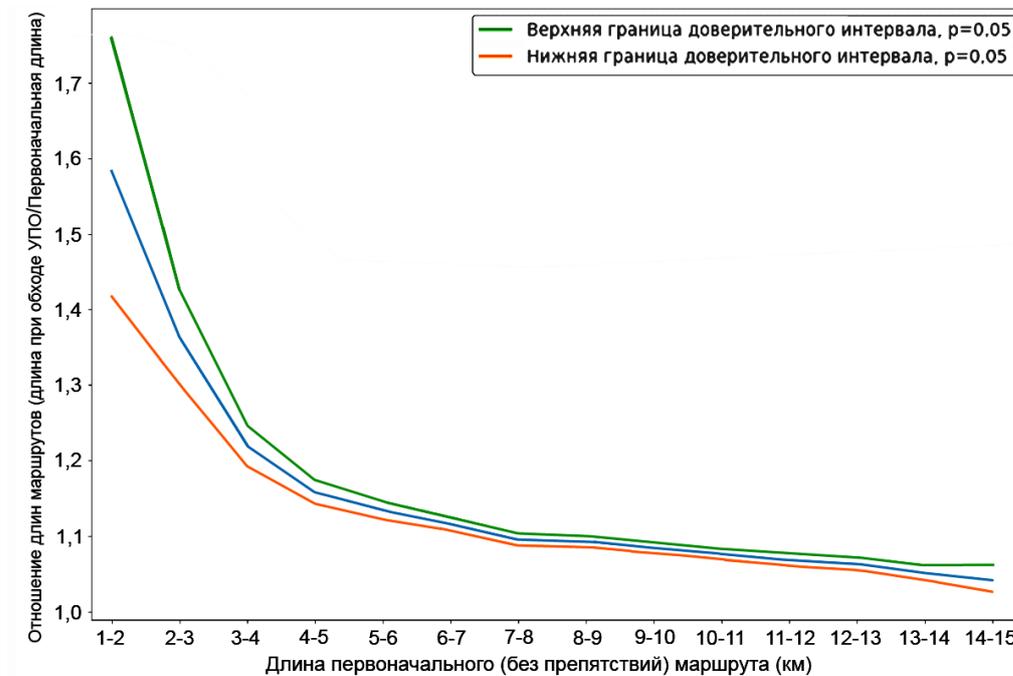


Рис. 4. Зависимость средних отношений длин маршрутов (длина при обходе УПО/ Первоначальная длина) для различной длины оригинального, проложенного без учета УПО, маршрута

Наконец, в таблице 3 дана более подробная описательная статистика отношений (Длина при обходе УПО / Первоначальная длина) для четырехкилометровых интервалов первоначальной длины маршрута, из которой видно, что все показатели (среднее, медиана, стандартное отклонение, разница между максимальным и минимальным значением) плавно уменьшаются по мере увеличения длины оригинального маршрута.

⁹ Например, для отношений длин, чья описательная статистика показана в табл. 1, тест [6] `scipy.stats.normaltest()` дает значение `pvalue = 0.0`.

Таблица 2. Зависимость средних отношений длин маршрутов (Длина при обходе УПО/Первоначальная длина) и числа вершин вдоль маршрутов (Число вершин при обходе УПО/Число вершин первоначального маршрута) для различной длины оригинального (проложенного без учета УПО) маршрута. Для отношений длин и числа вершин в скобках указаны 95 % доверительные интервалы

Расстояние (км)	Число маршрутов	Среднее отношение длин	Среднее отношение числа вершин
1-2	63	1,58 (1,404 – 1,747)	1,52 (1,33 – 1,69)
2-3	156	1,36 (1,299 – 1,423)	1,25 (1,17 – 1,32)
3-4	354	1,22 (1,192 – 1,245)	1,18 (1,13 – 1,22)
4-5	471	1,16 (1,142 – 1,174)	1,09 (1,06 – 1,11)
5-6	562	1,13 (1,122 – 1,146)	1,07 (1,04 – 1,09)
6-7	659	1,12 (1,107 – 1,125)	1,02 (1,00 – 1,04)
7-8	626	1,10 (1,088 – 1,103)	1,02 (1,01 – 1,04)
8-9	593	1,09 (1,085 – 1,100)	1,02 (1,00 – 1,03)
9-10	502	1,08 (1,077 – 1,091)	1,00 (0,98 – 1,02)
10-11	349	1,07 (1,069 – 1,083)	1,00 (0,97 – 1,02)
11-12	220	1,07 (1,060 – 1,076)	1,00 (0,96 – 1,03)
12-13	119	1,06 (1,054 – 1,072)	1,01 (0,97 – 1,04)
13-14	45	1,05 (1,040 – 1,061)	1,05 (1,00 – 1,09)
14-15	9	1,04 (1,021 – 1,058)	1,05 (1,01 – 1,08)

Таблица 3. Описательная статистика отношений (Длина при обходе УПО/Первоначальная длина) для выбранных интервалов первоначальной длины маршрута. Для среднего, медианы и стандартного отклонения в скобках указаны 95 % доверительные интервалы

Длина (км)	N	Среднее	Медиана	Стандартное отклонение	Мин.	Макс.	Диапазон Макс.–Мин.
1-5	1044	1,23 (1,216 – 1,255)	1,12 (1,106 – 1,137)	0,317 (0,286 – 0,358)	1,001	3,521	2,52
5-9	2440	1,11 (1,105 – 1,114)	1,07 (1,061 – 1,072)	0,115 (0,102 – 0,121)	1,000	1,707	0,707
9-13	1190	1,08 (1,073 – 1,081)	1,06 (1,060 – 1,067)	0,069 (0,067 – 0,077)	1,000	1,439	0,439
13-15	54	1,05 (1,04 – 1,059)	1,04 (1,028 – 1,045)	0,034 (0,027 – 0,043)	1,005	1,151	0,146

5. ОТНОСИТЕЛЬНЫЙ РИСК ДТП ПРИ ВЫБОРЕ МАРШРУТА, ОБХОДЯЩЕГО УПО

Как видно из предыдущего раздела, обход УПО гарантирует увеличение длины маршрута. При этом совсем не очевидно, что альтернативный маршрут будет более безопасным. Весьма важно поэтому оценить относительный риск ДТП при выборе маршрута, обходящего УПО.

Пусть необходимо проложить маршрут из A в B . Обозначим $A \rightarrow B$ маршрут с обходом УПО, а $A \rightarrow B$ — маршрут без обхода УПО. Пусть $ДТП1$ — число серьезных ДТП, зарегистрированных вдоль маршрута $A \rightarrow B$, $ДТП2$ — число серьезных ДТП вдоль маршрута $A \rightarrow B$. Тогда отношение $ДТП1/ДТП2$ будет мерой относительного риска (или отношением вероятностей того, что на маршруте произошло одно ДТП) на двух маршрутах (в предположении, что весь трафик перенаправляется с маршрута $A \rightarrow B$ на маршрут $A \rightarrow B$)¹⁰.

Для вычисления относительного риска нам, прежде всего, необходимо перебрать координаты всех серьезных ДТП, случившихся в Спрингфилде в 2013–2018 годах, и определить для каждого ДТП ближайшую к нему грань в графе G_0 . Для этого используется метод `get_nearest_edge()` пакета `OSMnx`.

К сожалению, `get_nearest_edge()` дает однозначные результаты лишь для однонаправленных граней и не в состоянии определить, какому направлению двунаправленной грани принадлежит конкретное ДТП. Чтобы узнать это, можно, казалось бы, сопоставить данные о направлении ДТП с данными о направлении трассы (например S, W, N, E), которой принадлежит данное ребро графа. Но такие данные известны далеко не для всех ДТП, к тому же, они не особо надежны и для трассы, идущей с севера на юг, может быть указано направление W или E.

Таким образом, нам остается только присвоить каждому направлению двунаправленной грани половину всех найденных ДТП и надеяться, что усреднение по многим граням для достаточно протяженного маршрута компенсирует неточность определения числа ДТП для конкретного направления двунаправленной грани.

Итак, с учетом замечаний о числе ДТП для двунаправленной грани подсчитаем для каждой пары маршрутов $A \rightarrow B$ и $A \rightarrow B$ число ДТП вдоль маршрута, после чего можно получить общую описательную статистику отношений $ДТП1/ДТП2$ по всем маршрутам (таблица 4).

Таблица 4. Общая статистика отношений длин маршрутов (ДТП вдоль $A \rightarrow B$ / ДТП вдоль $A \rightarrow B$)

Показатель	Значение	95 % Доверительный интервал
Среднее	0,91	(0,89 – 0,93)
Медиана	0,84	(0,83 – 0,842)
Стандартное отклонение	0,73	(0,49 – 0,96)
Минимум	0,13	
Максимум	27,75	

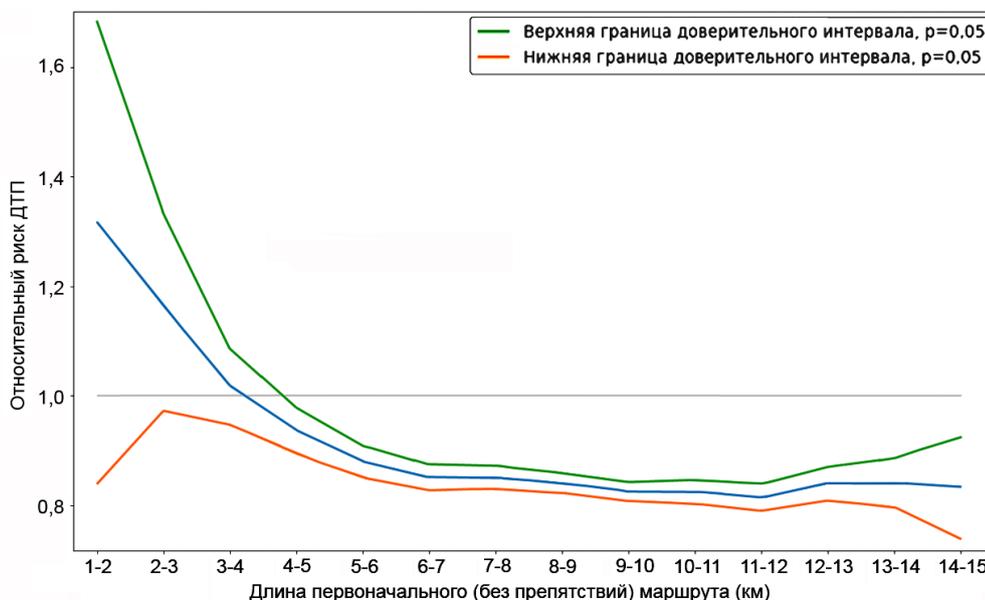
В таблице 5 показан относительный риск ДТП в зависимости от длины оригинального маршрута $A \rightarrow B$. Как и ранее, для вычисления доверительного интервала использовался бутстреп-метод.

Те же, что и в таблице 5, результаты более наглядно показаны на рис. 5, где серая горизонтальная прямая соответствует уровню относительного риска 1,0.

¹⁰ Здесь необходимо предположить, что выбор альтернативного маршрута происходит достаточно редко, иначе исказилась бы статическая картина ДТП, построенная без учета обхода препятствий.

Таблица 5. Зависимость средних отношений (ДТП вдоль $A \rightarrow B$ / ДТП вдоль $A \rightarrow B$) для различной длины маршрута $A \rightarrow B$

Расстояние (км)	N	Среднее отношение числа ДТП	95 % доверительный интервал
1-2	63	1,32	(0,834 – 1,689)
2-3	156	1,16	(0,978 – 1,335)
3-4	354	1,02	(0,948 – 1,087)
4-5	471	0,94	(0,894 – 0,979)
5-6	562	0,88	(0,850 – 0,909)
6-7	659	0,85	(0,828 – 0,874)
7-8	626	0,85	(0,830 – 0,871)
8-9	593	0,84	(0,823 – 0,859)
9-10	502	0,82	(0,808 – 0,842)
10-11	349	0,82	(0,803 – 0,846)
11-12	220	0,81	(0,789 – 0,839)
12-13	119	0,84	(0,809 – 0,870)
13-14	45	0,84	(0,796 – 0,886)
14-15	9	0,83	(0,738 – 0,926)

**Рис. 5.** Зависимость средних отношений (ДТП вдоль $A \rightarrow B$ / ДТП вдоль $A \rightarrow B$) для различной длины маршрута $A \rightarrow B$

Как видим, снижение относительного риска происходит только для достаточно протяженных маршрутов (в нашем случае — более 4 км). Что же касается меньших расстояний, то там, согласно таблице 2, происходит значительное увеличение протяженности альтернативного (обходящего УПО) маршрута и, как следствие, — увеличение числа ДТП вдоль него.

Наконец, в таблице 6 дана более подробная описательная статистика отношений (ДТП вдоль $A \rightarrow B$ / ДТП вдоль $A \rightarrow B$) для четырехкилометровых интервалов первоначальной длины маршрута. Как видим, все статистические показатели либо уменьшаются по мере увеличения первоначальной длины маршрута (стандартное

отклонение, разница между максимальным и минимальным значением), либо выходят на «плато» (среднее, медиана).

Таблица 6. Описательная статистика отношений (ДТП вдоль $A \rightarrow B$ /ДТП вдоль $A \rightarrow B$) для выбранных интервалов первоначальной длины маршрута. Для среднего, медианы и стандартного отклонения в скобках указаны 95 % доверительные интервалы

Длина (км)	N	Среднее	Медиана	Стандартное отклонение	Мин.	Макс.	Диапазон Макс.–Мин.
1–5	1044	1,02 (0,97–1,07)	0,85 (0,84–0,87)	0,80 (0,64–0,95)	0,17	10,0	9,83
5–9	2440	0,86 (0,84–0,87)	0,83 (0,82–0,84)	0,29 (0,27–0,31)	0,25	3,57	3,32
9–13	1190	0,82 (0,81–0,840)	0,84 (0,82–0,21)	0,20	0,27	1,56	1,29
13–15	54	0,83 (0,80–0,88)	0,85 (0,83–0,9)	0,15 (0,12–0,18)	0,43	1,22	0,79

Комбинируя таблицы 3 и 6, можно сопоставить увеличение среднего отношения длин маршрутов с изменением среднего относительного риска ДТП для различных интервалов протяженности оригинального (без учета УПО) маршрута (таблица 7). Так, например, для интервала 5–9 км имеем в среднем увеличение длины маршрута на 11 % и уменьшение относительного риска ДТП на 14 %.

Таблица 7. Зависимость изменения среднего относительного риска ДТП от среднего увеличения длины маршрута, обходящего УПО, для различных интервалов оригинального, не учитывающего УПО, маршрута. В скобках указаны соответствующие 95 % доверительные интервалы

Длина оригинального маршрута (км)	Увеличение среднего отношения длин маршрутов	Изменение среднего относительного риска ДТП
1–5	23% (21,6–25,5)%	+2% (–3%, +7%)
5–9	11% (10,5–11,4)%	–14% (–13%, –16%)
9–13	8% (7,3–8,1)%	–18% (–16%, –19%)
13–15	5% (4,1–5,9)%	–17% (–12%, –20%)

Для оригинальных (не учитывающих УПО) длин маршрута в интервале 1–5 км маршрут, обходящий УПО, в среднем на 23 % (95 % доверительный интервал: 21,6–25,5 %) длиннее, чем маршрут, не учитывающий препятствий, и риск ДТП возрастает в среднем на 2 % (95 % доверительный интервал: –3 %, +7 %).

Для оригинальных (не учитывающих УПО) длин маршрута в интервале 5–9 км маршрут, обходящий УПО, в среднем на 11 % (95 % доверительный интервал: 10,5–11,4 %) длиннее, чем маршрут, не учитывающий препятствий и риск ДТП уменьшается в среднем на 14% (95% доверительный интервал: 13–16 %).

Для оригинальных (не учитывающих УПО) длин маршрута в интервале 9–13 км маршрут, обходящий УПО, в среднем на 8 % (95 % доверительный интервал: 7,3–8,1 %) длиннее, чем маршрут, не учитывающий препятствий и риск ДТП уменьшается в среднем на 18 % (95 % доверительный интервал: 16–19 %).

Для оригинальных (не учитывающих УПО) длин маршрута в интервале 13–15 км маршрут, обходящий УПО, в среднем на 5 % (95 % доверительный интервал: 4,1–5,9 %)

длиннее, чем маршрут, не учитывающий препятствий и риск ДТП уменьшается в среднем на 17 % (95 % доверительный интервал: 12–20 %).

6. ВЫВОДЫ

Метакластеры серьезных ДТП, выявленные в предыдущей работе авторов [1], могут быть использованы соответствующими службами для выявления причин, приводящих к возникновению ДТП, и устранения их.

Водителям автомобилей, такси и автобусов в городе Спрингфилд при использовании маршрута протяженностью более 4 км можно рекомендовать алгоритм, позволяющий обойти метакластеры серьезных ДТП, описанный в данной статье, и тем самым снизить риск серьезного ДТП в среднем на 16 % при увеличении длины маршрута в среднем на 8 %. Авторы планируют применить подход, изложенный в данной статье, к российским мегаполисам, таким как Москва и Санкт-Петербург.

Список литературы

1. Герштейн А. М., Терехов А. Н. Выявление участков повышенной опасности на дорогах Массачусетса в 2013-2018 годах // Компьютерные инструменты в образовании. 2021. № 1. С. 27–40, doi: 10.32603/2071-2340-2021-1-27-40
2. Sahnoun I., Shawky M., Al-Ghafli A. Integrating Traffic Safety in Vehicle Routing Solution // Advances in Human Aspects of Transportation. AHFE 2017. Advances in Intelligent Systems and Computing. 2018. Vol. 597. P. 251–263. doi: 10.1007/978-3-319-60441-1_25
3. Graser A., Straub M., Dragaschnig M. Is OSM Good Enough for Vehicle Routing? A Study Comparing Street Networks in Vienna // Progress in Location-Based Services 2014. Cham, Germany: Springer, 2014. P. 3–17. doi: 10.1007/978-3-319-11879-6_1
4. Boeing G. OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks // Computers, Environment and Urban Systems. 2017. Vol. 65. P. 126–139. doi: 10.1016/j.compenvurbsys.2017.05.004
5. Agresti A., Franklin Ch., Klingenberg B. Statistics: The Art and Science of Learning from Data. 4th ed., Harlow, England: Pearson, 2018.
6. D'Agostino R., Pearson E. S. Tests for departure from normality. Empirical results for the distributions of b^2 and $\sqrt{b^1}$ // Biometrika. 1973. Vol. 60, № 3. P. 613–622. doi: 10.1093/biomet/60.3.613
7. Dijkstra E. W. A note on two problems in connexion with graphs // Numer. Math. 1959. Vol. 1. P. 269–271. doi: 10.1007/BF01386390
8. Bellman R. On a Routing Problem // Quarterly of Applied Mathematics. 1958. Vol. 16, № 1. P. 87–90. doi: 10.1090/QAM/102435
9. Huber P. J., Ronchetti M. E. 7.7 Concomitant scale estimates // Robust Statistics. Hoboken, NJ, U.S.: Wiley, 2009, P. 172–174.
10. Owen A. B. A Robust Hybrid of Ridge and Lasso and Ridge Regression, 2006. URL: <https://statweb.stanford.edu/~owen/reports/hhu.pdf> (date: 23.06.2022).
11. Koenker R., Hallock K. F. Quantile Regression // Journal of Economic Perspectives. 2001. Vol. 15, № 4. P. 143–156.

Поступила в редакцию 10.06.2022, окончательный вариант — 23.06.2022.

Герштейн Аркадий Михайлович, аспирант математико-механического факультета СПбГУ,
✉ ArkadyGer@gmail.com

Терехов Андрей Николаевич, доктор физико-математических наук, профессор, заведующий кафедрой системного программирования математико-механического факультета СПбГУ, a.terekhov@spbu.ru

Computer tools in education, 2022

№ 2: 5–18

<http://cte.eltech.ru>

[doi:10.32603/2071-2340-2022-2-5-18](https://doi.org/10.32603/2071-2340-2022-2-5-18)

Routing Algorithm for Vehicles that Avoids Severe Traffic Accident Hotspots on the Road Network (Using the City of Springfield, Massachusetts as a Case Study)

Gershtein A. M.¹, Postgraduate, ✉ ArkadyGer@gmail.com
Terekhov A. N.¹, PhD, Professor, a.terekhov@spbu.ru

¹Saint Petersburg State University, 7–9, Universitetskaya emb., 198504, Saint Petersburg, Starii Petergof, Russia

Abstract

In [1] statistically significant clusters (hotspots) of severe Traffic Accidents (TA) are found. In this article, as a continuation of [1], a simple routing algorithm to avoid TA hotspots on a road network has been proposed («hotspot avoidance» path). If the road network is represented by a graph with edges and nodes, it is enough to mark every edge which lead to the TA hotspot as «not passable» by letting a <Length> attribute of the edge be a very large digit, much greater than max edge length for a given road graph — and the routing algorithm (Dijkstra or Bellman-Ford) will avoid the TA hotspot automatically.

Computer simulation was performed for Springfield, MA. It is shown that for the same initial and end points of the route, an average ratio (Route avoiding TA length/Original route length) is bigger for shorter original (without taking into account TA hotspots) routes and gradually slows down to 1.04 for max original route length inside Springfield.

Route length ratios show extra route length needed to avoid TA hotspots, but say nothing about new route safety. To estimate safety gain, a new Relative Risk Ratio RRR= (TAs along route which avoids TA hotspots/TAs along original route) was introduced.

It is shown for Springfield that relatively short (less than 4 km) original routes are more dangerous (have more TAs along the «hotspot avoidance» route) than original ones, but for relatively long (> 4 km) original routes average RRR gets smaller by 16 % while modified path gets longer by 8 % in average.

Keywords: *routing vehicle traffic, relative risk ratio, Dijkstra algorithm, accident hotspot, cluster, DBSCAN, Monte-Carlo simulation, Massachusetts.*

Citation: A. M. Gershteyn and A. N. Terekhov, “Routing Algorithm for Vehicles that Avoids Severe Traffic Accident Hotspots on the Road Network (Using the City of Springfield, Massachusetts as a Case Study),” *Computer tools in education*, no. 2, pp. 5–18, 2022 (in Russian); doi: 10.32603/2071-2340-2022-2-5-18

References

1. A. M. Gershteyn and A. N. Terekhov, “Hotspots of Traffic Accident that cause injuries or death in Massachusetts from 2013 to 2018,” *Computer tools in education*, no. 1, pp. 45–57, 2021 (in Russian); doi: 10.32603/2071-2340-2021-1-46-58
2. I. Sahnoun, M. Shawky, and A. Al-Ghafli, “Integrating Traffic Safety in Vehicle Routing Solution,” in *Advances in Human Aspects of Transportation. AHFE 2017. Advances in Intelligent Systems and Computing*, vol. 597, pp. 251–263, 2018; doi: 10.1007/978-3-319-60441-1_25

3. A. Graser, M. Straub, and M. Dragaschnig, "Is OSM Good Enough for Vehicle Routing? A Study Comparing Street Networks in Vienna," in *Progress in Location-Based Services 2014*, Cham, Germany: Springer, 2014, pp. 3–17; doi: 10.1007/978-3-319-11879-6_1
4. G. Boeing, "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *Computers, Environment and Urban Systems*, vol. 65, pp. 126–139, 2017; doi: 10.1016/j.compenvurbsys.2017.05.004
5. A. Agresti, Ch. Franklin, and B. Klingenberg, *Statistics: The Art and Science of Learning from Data*, 4th ed., Harlow, England: Pearson, 2018.
6. R. D'Agostino, and E. S. Pearson, "Tests for departure from normality. Empirical results for the distributions of b^2 and $\sqrt{b^1}$," *Biometrika*, vol. 60, no. 3, pp. 613–622, 1973; doi: 10.1093/biomet/60.3.613
7. E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.* vol. 1, pp. 269–271, 1959; doi: 10.1007/BF01386390
8. R. Bellman, "On a Routing Problem," *Quarterly of Applied Mathematics*, vol. 16, no. 1. pp. 87–90, 1958; doi: 10.1090/QAM/102435
9. P. J. Huber and M. E. Ronchetti, "7.7 Concomitant scale estimates," in *Robust Statistics*, Hoboken, NJ, U.S.: Wiley, 2009, pp. 172–174.
10. A. B. Owen, "A Robust Hybrid of Ridge and Lasso and Ridge Regression," in *Statweb.stanford.edu*, 2006. [Online]. Available: <https://statweb.stanford.edu/~owen/reports/hhu.pdf>
11. R. Koenker and K. F. Hallock, "Quantile Regression," *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 143–156, 2001.

Received 10-06-2022, the final version — 23-06-2022.

Arkadiy Gershtein, Postgraduate of the Faculty of Mathematics and Mechanics, SPbSU,
✉ ArkadyGer@gmail.com

Andrey Terekhov, PhD, Professor, Head of the Department of System Programming of the Faculty of Mathematics and Mechanics, SPbSU, a.terekhov@spbu.ru