

Анализ данных для прогнозирования вероятности дорожно-транспортных происшествий с участием пешеходов

Е.А. Кузьменко, Д.С. Донченко, В.О. Рагозин

Волгоградский Государственный Технический Университет

Аннотация: Статья посвящена сбору и обработке данных для обучения модели прогнозирования вероятности дорожно-транспортных происшествий с участием пешеходов. Для выявления факторов, влияющих на ДТП, был приведен анализ статистической информации, взятой из «Портала Открытых Данных Правительства Москвы» и портала «БезопасныеДороги.Рф», а также с карточек ДТП сайта ГИБДД.

Ключевые слова: дорожно-транспортное происшествие, анализ данных, безопасность пешеходов, машинное обучение, градиентный бустинг на решающих деревьях.

Введение

При создании проектов по организации дорожного движения первым приоритетом должна стоять безопасность всех его участников. До настоящего времени потребностям самых уязвимых участников дорожного движения – пешеходам в России не уделялось должного внимания. Анализ аварийности на автомобильном транспорте показывает, что одним из наиболее распространённых видов ДТП в городах России являются наезды на пешеходов, характеризующиеся высокой тяжестью последствий, т.е. наезды ежегодно составляют 30-50 % от всех ДТП с пострадавшими (рис. 1) [1].



Рис. 1. – Статистика ДТП с участием пешеходов за 2019 год

Наиболее безопасными, по статистике, являются дороги стран Европы, где этот показатель не превышает 10 человек на 100 тыс. населения [2]. В этих странах действуют организации, принимающие меры по повышению безопасности для пешеходов: Vision Zero в Швеции, TRAFI в Финляндии, POBR в Польше, реализуют проекты по автоматизированному сбору и анализу данных о ДТП [3].



Рис. 2. – Статистика смертности в ДТП за 2018 год

В России исследованию проблем безопасности пешеходов уделяется мало внимания. Поэтому для выработки обоснованных решений по организации дорожного движения необходимо использовать зарубежный опыт и учитывать местные условия и особенности.

Модель прогнозирования ДТП

Для предотвращения возникновения ДТП предлагается использовать результаты прогнозирования аварий с участием пешеходов на особо опасных участках дорог. В качестве метода прогнозирования будет использоваться градиентный бустинг на решающих деревьях.

Градиентный бустинг на решающих деревьях на сегодняшний день является одним из самых мощных алгоритмов прогнозирования. Это достигается благодаря адаптивной технике построения композиции. Градиентный бустинг представляет собой алгоритм, построенный на простых предсказателях (деревьях решений), которые улучшают целевую функцию [4].

В качестве инструмента для построения модели будет использоваться CatBoost – библиотека для градиентного бустинга на решающих деревьях. В ней поддержано несколько видов деревьев, в том числе oblivious (симметричные) деревья, которые используются по умолчанию. Они быстро учатся и помогают обучению быть более устойчивым к изменению параметров с точки зрения изменений итогового качества модели, что сильно уменьшает необходимость в подборе параметров. Для предотвращения переобучения модели CatBoost имеет собственную реализацию алгоритма градиентного бустинга, которая уменьшает смещения остатков [5].

Тем самым, CatBoost имеет преимущества перед аналогами: он устойчив к переобучению, может работать с категориальными признаками без дополнительной предобработки, то есть не требует преобразования набора данных в какой-либо определенный формат, как XGBoost и LightGBM [6].

Перед обучением модели нужно подготовить обучающий набор, так как готовых с нужными и полными данными по факторам ДТП нет.

Выделение факторов, влияющих на риск возникновения ДТП

Для правильного планирования и организации мер по повышению безопасности дорожного движения важно выделить основные факторы, способствующие возникновению ДТП.

Поскольку в дорожно-транспортных происшествиях участвуют многие переменные из трех основных групп, таких как человеческий фактор, факторы транспортного средства и факторы дорожной среды это усложняет подготовку модели множественной регрессии. Факторы риска могут проявляться в различных сочетаниях [7-8].

Так как модель разрабатывается для проектировщика по организации дорожного движения, то целевым объектом будет сфера его влияния – дорожная инфраструктура.

Из группы внешних факторов были выбраны только те переменные, которые можно изменить с помощью реорганизации дорожного движения, или влияние которых снизится после внесения изменений в дорожную среду. Например, пасмурная погода или темное время суток перестают иметь значение, когда улица хорошо освещена.

Поэтому важно обладать информацией обо всех факторах влияния, как о главных, так и о неблагоприятных внешних, помимо сведений, указанных в учетной карточке ДТП, которые зачастую бывают неполными или недостаточно точными. Зная привязку к километражу дороги, где исторически случались ДТП, согласно отметкам в учетных карточках, необходимо дополнять их факторами, сопутствовавшими ему, которые дают возможность установить причину ДТП. Точное выявление факторов – основа выбора правильного решения для повышения безопасности дорожного движения на конкретном участке дорожной сети. Они не всегда бывают очевидными, но имеют логическое обоснование – в некоторых случаях неверно расположенные объекты дорожной инфраструктуры мешают переходу дороги, удлиняют путь или являются неудобными для определенных категорий пешеходов (с колясками, лицам преклонного возраста, тем, кому сложно спускаться и подниматься по ступеням), что может побуждать нарушать правила дорожного движения и пересекать проезжую часть в неполюженном месте.

В качестве входных данных использовались: широта и долгота, состояние погоды (может указать, что на дороге в определенных погодных условиях появляются травмоопасные факторы, такие, как недостаточное освещение, плохое качество дорожного покрытия, которые могут быть напрямую не указаны в карточке ДТП или неактуальны в получаемых сведениях), состояние проезжей части, освещение, нарушение ПДД (в зависимости от типа нарушения можно изучать, что к этому привело),

объекты улично-дорожной сети (УДС) на месте происшествия и вблизи него, непосредственные нарушения со стороны водителя и пешехода, недостатки транспортно-эксплуатационного содержания УДС, дорожные знаки, дорожная разметка, пешеходных мосты, подземные переходы, выходы из метрополитена, количество полос, максимально допустимая скорость. В качестве выходных данных – вероятность ДТП и степень тяжести последствий.

Сбор данных

Были выгружены карточки ДТП с участием пешеходов за 2016-2019 года с сайта ГИБДД. Карточки представлены в формате excel-файла на двух страницах [1].

С помощью библиотеки `openruxl Python` из них были извлечены и добавлены в csv файл данные о: времени, широте, долготе, состоянии погоды, состоянии проезжей части, освещении, нарушениях ПДД, степени тяжести последствий, объектах улично-дорожной сети (УДС) на месте происшествия и вблизи него, непосредственных нарушениях со стороны водителя и пешехода, недостатках транспортно-эксплуатационного содержания УДС [1].

Далее, были выгружены данные с сайта Портал Открытых Данных Правительства Москвы [9] и БезопасныеДороги.Рф [10]: дорожные знаки, дорожная разметка, пешеходных мосты, подземные переходы, выходы из метрополитена в формате csv (рис. 3) [9]. Информация из них была добавлена в общий csv-файл с помощью функций модуля `csv Python`, поиск данных велся в пределах 50 метров от случая ДТП, т.к. координаты могут быть неточными.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	AdmArea	District	Location	Status	HaveSign	TrafficLight	TLComme	Longitude	Latitude	global_id	geoData					
2	Централь	район Яки	Лаврушин	Эксплуатс	да			37.619152	55.743923	9,34E+08	{ "coordinates": [37.619152, 55.743923], "type": "Point" }					
3	Южный а,	Донской	Донская у	Эксплуатс	да			37.599683	55.714218	9,34E+08	{ "coordinates": [37.599683, 55.714218], "type": "Point" }					
4	Северо-В	район Ма	Веткина у	Эксплуатс	да			37.615519	55.806922	9,34E+08	{ "coordinates": [37.61551903, 55.80692204], "type": "Point" }					
5	Централь	Мещански	Садовая- (Эксплуатс	да			37.618951	55.773852	9,34E+08	{ "coordinates": [37.618951, 55.773852], "type": "Point" }					
6	Западный	район Раг	Ломоносс	Эксплуатс	да			37.513861	55.706058	9,34E+08	{ "coordinates": [37.513861, 55.706058], "type": "Point" }					
7	Западный	район До	Бережко	Эксплуатс	да			37.552155	55.731061	9,34E+08	{ "coordinates": [37.552155, 55.731061], "type": "Point" }					
8	Восточны	район Сои	Щербакoi	Эксплуатс	да			37.722502	55.782335	9,34E+08	{ "coordinates": [37.722502, 55.782335], "type": "Point" }					
9	Северо-В	район Ма	Шеремет	Эксплуатс	да			37.615707	55.803603	9,34E+08	{ "coordinates": [37.615707, 55.803603], "type": "Point" }					
10	Централь	район Ари	Волхонка	Эксплуатс	да			37.608798	55.749187	9,34E+08	{ "coordinates": [37.608798, 55.749187], "type": "Point" }					
11	Централь	Тверской	Боровици	Эксплуатс	да			37.610753	55.748784	9,34E+08	{ "coordinates": [37.610753, 55.748784], "type": "Point" }					
12	Централь	район Хаг	Пречисте	Эксплуатс	да			37.586496	55.738602	9,34E+08	{ "coordinates": [37.58649644, 55.73860261], "type": "Point" }					
13	Централь	район Хаг	Комсомо	Эксплуатс	да			37.594399	55.735247	9,34E+08	{ "coordinates": [37.59439945, 55.73524752], "type": "Point" }					
14	Централь	Мещански	Гиляровс	Эксплуатс	да			37.633385	55.790093	9,34E+08	{ "coordinates": [37.63338522, 55.79009332], "type": "Point" }					
15	Централь	Тверской	Моховая	Эксплуатс	да			37.615114	55.756925	9,34E+08	{ "coordinates": [37.615114, 55.756925], "type": "Point" }					
16	Южный а,	Даниловс	Андропо	Реконстр	да			37.666444	55.696361	9,34E+08	{ "coordinates": [37.666444, 55.696361], "type": "Point" }					
17	Централь	Тверской	Тверской	Эксплуатс	да			37.604931	55.765021	9,34E+08	{ "coordinates": [37.604931, 55.765021], "type": "Point" }					
18	Западный	район Раг	Джавахар	Эксплуатс	да			37.536127	55.693325	9,34E+08	{ "coordinates": [37.536127, 55.693325], "type": "Point" }					
19	Северо-В	район Ма	Шеремет	Эксплуатс	да			37.615946	55.798449	9,34E+08	{ "coordinates": [37.61594627, 55.7984499], "type": "Point" }					
20	Северо-В	район Би	Плещев	Эксплуатс	да			37.605741	55.884254	9,34E+08	{ "coordinates": [37.605741, 55.884254], "type": "Point" }					
21	Юго-Запа	Академи	Профсою	Эксплуатс	да			37.566568	55.681310	9,34E+08	{ "coordinates": [37.56656885, 55.68131069], "type": "Point" }					
22	Восточны	район Ив	Энтузиаст	Эксплуатс	да			37.82016	55.771573	9,34E+08	{ "coordinates": [37.82016, 55.771573], "type": "Point" }					
23	Централь	Пресненс	Брестская	Эксплуатс	да			37.588692	55.772011	9,34E+08	{ "coordinates": [37.58869287, 55.77201168], "type": "Point" }					

Рис.3 – Пример структуры csv-файла о выходах из метрополитена

Выводы

В статье обоснована необходимость использования прогнозных моделей для принятия обоснованных решений по снижению рисков ДТП с участием пешеходов. Проведен анализ данных, влияющих на аварийность. Выявлены факторы, которые могут быть изменены под воздействием управляющих воздействий и могут повлиять на снижение ДТП с участием пешеходов. Рассмотрены вопросы сбора и обработки данных для построения прогнозной модели.

Литература

1. Статистика: Показатели состояния безопасности дорожного движения // Официальный сайт Госавтоинспекции. URL: stat.gibdd.ru (дата обращения: 19.04.20).
2. Global status report on road safety 2018 // World Health Organization. URL: who.int/publications/i/item/global-status-report-on-road-safety-2018 (accessed: 19.04.20).
3. Лучшие европейские практики в области профилактики безопасности дорожного движения // Общероссийская детско-молодежная общественная организация по развитию просветительских программ

«Здоровая Нация». URL: zdorn.ru/attachments/article/240/otchet.pdf (дата обращения: 19.04.20).

4. Соколов, Е. Семинары по композиционным методам // MachineLearning. 2015. URL: machinelearning.ru/wiki/images/7/7e/Sem03_ensembles_2014.pdf (дата обращения: 19.04.20).

5. Overview of Catboost // Catboost. URL: catboost.ai/docs/concepts/about.html (дата обращения: 19.04.20).

6. Dorogush, A.V., Ershov, V., Gulin, A. CatBoost: gradient boosting with categorical features support // Yandex. 2018. URL: arxiv.org/pdf/1810.11363.pdf (accessed: 19.04.20).

7. Капитанов, В.Т., Моница, О.Ю., Сильянов, В.В., Чубуков, А.В. Некоторые аспекты вероятностно-статистического моделирования аварийности в регионах Австрии, Германии, России // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2014. № 4 (39). С. 97-102.

8. Митник, В.М. Проблемы аварийности на дорогах и пути ее снижения // Вестник Самарской гуманитарной академии. Серия: Право. 2013. №1 (13). С. 110.

9. Портал открытых данных Правительства Москвы. URL: data.mos.ru (дата обращения: 19.04.20).

10. OpenStreetMap. URL: openstreetmap.org (дата обращения: 19.04.20).

References

1. Statistika: Pokazateli sostojanija bezopasnosti dorozhnogo dvizhenija [Statistics: road traffic safety indicators] Oficial'nyj sajt Gosavtoinspekcii. URL: stat.gibdd.ru.

2. Global status report on road safety 2018 World Health Organization. URL: who.int/publications/i/item/global-status-report-on-road-safety-2018 (accessed: 19.04.20).
3. Luchshie evropejskie praktiki v oblasti profilaktiki bezopasnosti dorozhnogo dvizhenija [The best European practices in the prevention of road traffic safety]. Obshherossijskaja detsko-molodezhnaja obshhestvennaja organizacija po razvitiyu prosvetitel'skih programm «Zdorovaja Nacija». URL: zdorn.ru/attachments/article/240/otchet.pdf.
4. Sokolov, E. Seminary po kompozicionnym metodam [Seminars on computer methods] MachineLearning. 2015. URL: machinelearning.ru/wiki/images/7/7e/Sem03_ensembles_2014.pdf.
5. Overview of Catboost. Catboost. URL: catboost.ai/docs/concepts/about.html (data obrashhenija: 19.04.20).
6. Dorogush, A.V., Ershov, V., Gulin, A. CatBoost: gradient boosting with categorical features support. Yandex. 2018. URL: arxiv.org/pdf/1810.11363.pdf (accessed: 19.04.20).
7. Kapitanov, V.T., Monina, O.Yu., Sil'yanov, V.V., Chubukov, A.V. Vestnik Moskovskogo avtomobil'no-dorozhnogo gosudarstvennogo tehničeskogo universiteta (MADI). 2014. № 4 (39). pp. 97-102.
8. Mitnik, V.M. Vestnik Samarskoj gumanitarnej akademii. Serija: Pravo. 2013. №1 (13). P. 110.
9. Portal Otkrytyh Danyh Pravitel'stva Moskvy [Open Data Portal of the Moscow Government]. URL: data.mos.ru.
10. OpenStreetMap. URL: openstreetmap.org.