

## **Использование методов обработки и анализа больших данных для обнаружения опасных участков дороги**

**Осипов А.В.**, к.физ-мат.наук, доцент кафедры «Прикладная информатика», Московский Политехнический университет, Москва, Россия

**Суворов С.В.**, к.э.н., профессор кафедры «Прикладная информатика», Московский Политехнический университет, Москва, Россия

**Жиляева И.А.**, к.э.н., доцент кафедры «Прикладная информатика», Московский Политехнический университет, Москва, Россия

**Шаруев Д.М.**, магистрант кафедры «Прикладная информатика», ФГБОУ ВО «Московский политехнический университет», Москва, Россия

**Аннотация.** В данной статье была рассмотрена возможность использования больших данных для предотвращения дорожно-транспортных происшествий. А также, рассмотрена возможность обнаружения опасных участков дороги с помощью нейронной сети.

**Ключевые слова:** ДТП, США, SPSS, статистика, большие данные, нейросеть, прогнозирование

### **Using Big Data Processing and Analysis Techniques to Prevent Accidents**

**Osipov A.V.**, candidate of physico-mathematical sciences, Associate Professor of the «Applied Informatics» Department at the Moscow Polytechnic University, Moscow, Russia

**Suvorov S.V.**, candidate of economic sciences, Associate Professor of the «Applied Informatics» Department at the Moscow polytechnic University, Moscow, Russia

**Zhilyaeva I.A.**, candidate of economic sciences, Associate Professor of the «Applied Informatics» Department at the Moscow polytechnic University, Moscow, Russia

**Sharuev D.M.**, magistrate of the «Applied Informatics» Department at the Moscow polytechnic University, Moscow, Russia

**Annotation.** This article discusses the possibility of using big data to prevent road accidents. Also, the possibility of detecting dangerous road sections using a neural network is considered.

**Keywords:** Accident, USA, SPSS, statistics, big data, neural network, forecasting

В современном мире мегаполисы движутся в направлении построения умной городской инфраструктуры. Российская Федерация не является исключением, примером является ведомственный проект Цифровизация городского хозяйства «Умный город» утвержденного приказом Министерства строительства и жилищно-коммунального хозяйства России от 31 октября 2018 года.

Кузнецова А.И. утверждает, что развитие городской инфраструктуры необходимо для увеличения уровня жизни населения, а также развития экономических показателей региона [1,3-5].

Авторами работы [2] предложена оптимизация потоков автономных транспортных средств в городских условиях. С помощью метода ближайших соседей, реализованного на платформе Hadoop-Spark, производится тактическое перераспределение потоков транспортных средств. Модель показала хорошие результаты при штатной работе системы. Однако, на дороге периодически возникают внештатные ситуации, что свидетельствует о том что модель может быть использована не в каждом случае, ввиду того что возникновение дорожно-транспортных происшествий не предусмотрено на опасных участках, таких как переезды, перекрёстки и крутые повороты.

Для нахождения закономерностей возникновения дорожно-транспортных происшествий, необходимо уделить особое внимание обработке массивов данных о ДТП, так как полный и всесторонний анализ данных является одним из важнейших аспектов принятия решений по улучшению ситуации на дорогах. Именно результаты такого анализа имеют важное значение для выработки

решений в области обеспечения безопасности дорожного движения, в том числе по совершенствованию его организации. В статье Д.Н. Полякова [3] отмечается, что безопасное дорожное движение характеризуется отсутствием не дорожно-транспортной аварийности, а опасных условий и действий (бездействия), приводящих к ДТП.

Куприяновский В.П. в своей статье цифровая трансформация экономики, железных дорог и умных городов, планы и опыт Великобритании заметил, что «Крупные инвестиции в настоящее время реализуются в Великобритании в программе интеллектуальных автострад, которая использует цифровые технологии, когда это необходимо для отслеживания инцидентов, передачи инструкции в режиме реального времени для водителей, и открытия резервных полос и снятия при необходимости ограничений. Данные из пилотных программ подтверждают результат в 22% по уменьшению времени в пути и сокращению числа инцидентов на 55,7%.»[4]

Несмотря на то, что каждое конкретное ДТП представляет собой случайное явление, статистический анализ большого объема информации позволяет находить общие закономерности их возникновения.

Для анализа были использованы данные о дорожно-транспортных происшествиях в Соединенных Штатах Америки за 2016-2018гг., предоставленные Sobhan Moosavi под лицензией Creative Commons Attribution-Noncommercial-ShareAlike license (CC BY-NC-SA 4.0). Дата сет насчитывает более 2 млн. строк и 49 переменных. [6,7]

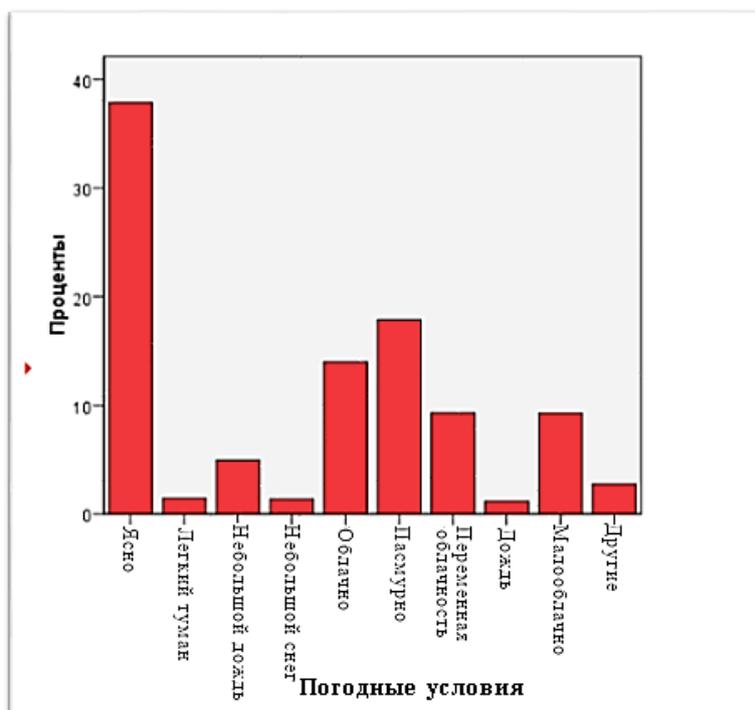
В исходных данных количество записей соответствует количеству происшествий, а данные в столбцах являются уточняющими: когда, где и при какой погоде произошло ДТП.

Рассмотрим количество произошедших инцидентов в зависимости от штата, в котором они произошли. Это необходимо для выявления зависимости между происшествием и его географическим положением.

Наибольшее количество происшествий происходит на восточном побережье Соединенных Штатов Америки, что может быть связано с тем что

плотность населения, а также количество дорожных развязок значительно превышает число на западе страны. Более того, Восточное побережье в связи с более мягким климатом привлекает большое количество туристов, которые могут не знать особенности дорожного движения в Америке, так же увеличивает общий процент инцидентов. Так же от штата к штату могут быть значительные изменения в ПДД.

Но не только плотность населения и туристические потоки влияют на ситуацию на дороге. Немаловажным фактором являются погодные условия, которые были во время происшествия. На рис.1 представлен график, наглядно демонстрирующий в какую погоду, в процентном соотношении, произошло ДТП.



*Рис. 1. – Количество ДТП в зависимости от погоды*

Очевидно, что большая часть происшествий происходит в ясную, безоблачную погоду, в связи с особенностями климата США. На втором месте количество происшествий, произошедших в пасмурную погоду. Это может быть связано с повышением влажности воздуха, и, как факт, ухудшением

самочувствия у участников дорожного движения, что приводит к уменьшению скорости реакции на аварийную ситуацию.

После проведения анализа различных факторов, которые в большей или меньшей степени могут влиять на дорожные происшествия, целесообразно использовать другой подход к исследованию имеющихся данных. Одним из возможных решений является создание нейросети, которая будет предсказывать возможную гео-локацию будущих происшествий, учитывая погодные условия, серьезность аварии, время суток, и множество других факторов, представленных в наборе. В качестве инструмента для создания нейросети была выбрана аналитическая платформа Deductor Pro, позволяющая решать задачи кластеризации и прогнозирования.

Для наглядного представления результатов на рис. 2 показано вероятное происшествие, произошедшее на железнодорожных путях, отмеченное маркером. А также предсказанное нейросетью место происшествия с использованием координат полученных в результате работы алгоритма.



***Рис. 2 – Возможное место происшествия, предсказанное нейронной сетью***

На данном примере, представленном на рисунке, видно, что прогнозируемое значение расходится с местом повышенной опасности на несколько метров, что является достаточно точным результатом, из чего можно

сделать вывод о целесообразности использования нейросети в качестве инструмента предотвращения будущих ДТП на основе данных о прошлом.

Но применим ли полученный алгоритм в других географических зонах? Мы предлагаем провести исследования дорожной ситуации в России. Для этого, проведем анализ двух стран, а именно России и США. По данным ГИБДД Российской Федерации 77% ДТП происходят в городах. Согласно отчету, за 2016 год по США: Количество ДТП в городах было больше, чем число ДТП в сельской местности в 2016 и 2017 годах и составляет 59% и 41% соответственно.

Зарегистрированных водителей в РФ – около 42% населения (2010 г.), то есть (с учетом роста за последние годы) около 50%, что в абсолютных числах составляет примерно 73 млн. По США есть более точная статистика – 221,712 млн. С правами на вождение автотранспорта в США 68,8% населения.

Согласно данным ГИБДД в России, зарегистрировано 60,5 млн транспортных средств, из них 20,9 млн старше 15 лет, еще 10 млн старше десяти лет, что свидетельствует о том, что около половины автомобилей из-за их состояния могут являться причиной аварийных ситуаций, в США (2016) – 288,034 млн. То есть зарегистрированных авто в США в 5,33 раза больше и качество автотранспорта значительно выше.

Общая протяженность дорожной сети: РФ – 1 507 750 км, США – 6 733 024 км. За прошедшие пару лет в России прибавилось порядка 100 тысяч километров дорог (было 1396000). Однако, при вдвое меньшей площади, дорог в США в 4,47 раза больше, чем в России. Для сравнения в Канаде 1420300 километров дорог. А в Китае – 4859500 км.

Далее разберем дорожную ситуацию в странах, на примере двух похожих городов, располагающихся недалеко от крупного города, развязки или шоссе, а также со схожим расположением транспортных линий: Жуковский (Московская область) и Спаркс (США, Невада). Данные города были выбраны не случайно, они находятся в двухчасовой доступности от главных городов тех регионов, в которых они находятся, из чего можно сделать вывод о том, что города схожи по географическому положению.

В городском округе Жуковский протяженность автомобильных дорог межмуниципального или регионального значения составляет 42,276 км, местного значения - 28,44 км. В Спаркс протяженность дорог составляет около 60 километров (37 миль). Численность населения обоих городов на 2016 год приблизительно равна 100 000 человек. Основные дорожные развязки находятся на северо-западе городов. На западе находятся междугородние трассы, и в том и в другом городе планируется создать развязку, которая улучшит ситуацию на дорогах. Карты городов представлены на рис. 3.



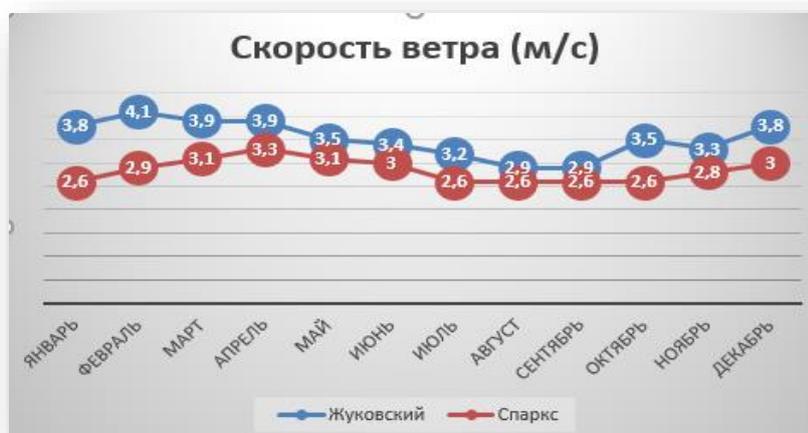
*Рис.3 – Карта городов Спаркс и Жуковский*

Рассмотрим такие климатические особенности данных городов, как температура воздуха, количество осадков, а также скорость ветра. Так как эти факторы могут влиять на шанс возникновения дорожно-транспортных происшествий. Сравнение климата городов представлено на рис. 4-6.

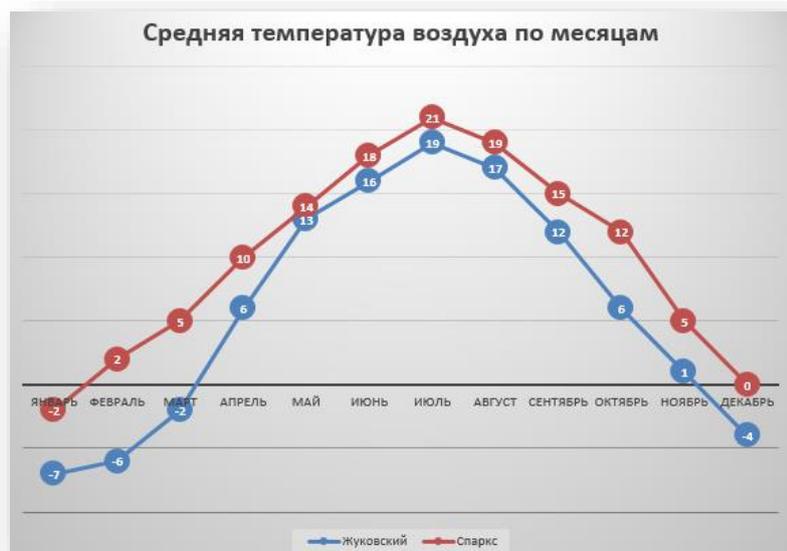
Приведенные факторы могут напрямую влиять на дорожную ситуацию, например, высокая температура воздуха пагубно отражается на самочувствии человека. Осадки могут создавать аварийную ситуацию посредством изменения свойств дорожного полотна.



*Рис. 4 – Количество осадков (мм)*



*Рис. 5 – Скорость ветра (м/с)*



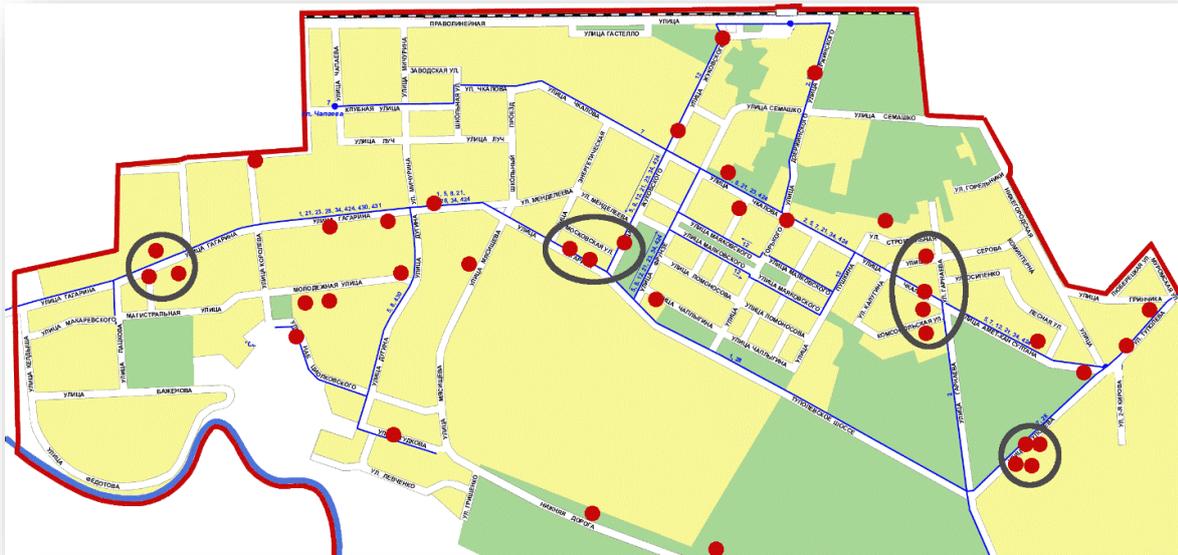
*Рис.6 – Средняя температура воздуха по месяцам (°C)*

На основе имеющихся данных о погодных условиях можно сделать вывод, что температура воздуха, а также скорость ветра в городах отличается незначительно. Тогда как среднее количество осадков, особенно в летние месяцы, сильно разнится, это связано с тем, что Спаркс находится южнее Жуковского и имеет более засушливый климат. Следовательно, можно сделать вывод о том, что города почти идентичны друг другу.

По данным статистики в городе Жуковский в 2016 году погибло 4 человека. [5] В Спаркс за тот же год погибло 7 человек. Чтобы избежать таких случаев в дальнейшем предлагается изучать статистические данные об произошедших ДТП. Однако такой большой объем информации невозможно обработать, используя лишь человеческий ресурс. В этом может помочь нейронная сеть, которая на основе внесенных в нее данных может заранее находить опасные участки на дороге. Полученные результаты можно будет использовать повсеместно, находя похожие друг на друга географические районы как было представлено выше.

Применим полученный алгоритм на город Жуковский с целью узнать будет ли алгоритм находить опасные участки дороги и там. Мы обратились в администрацию города Жуковский с целью получения данных по дорожно-транспортным происшествиям за 2016 год чтобы сравнить результаты полученные в ходе работы нейросети. На рис. 7 представлена карта города Жуковский с выделенными областями в которых ДТП происходили наиболее кучно.

В ходе работы алгоритма было получено большое количество результатов они будут представлены в виде тепловой карты, в которой красным цветом выделены места с наибольшим количеством точек, а синим с меньшим. Результаты представлены на рис. 8.



*Рис. 7 – Карта города Жуковский с ДТП*



*Рис. 8 – Карта попадания результатов работы нейросети*

Как можно заметить результаты полученные в ходе работы нейросети близко к реальным значениям, на пересечении трех улиц Чкалова, ул. Амет-Хан Султана и ул. Гарнаева нейросеть показала опасный участок дороги на котором за 2016 произошло 3 ДТП, 2 из которых наезд на пешехода. Это может быть связано с тем, что непосредственно перед пешеходными переходами,

отсутствуют лежащие полицейские. Этот перекресток с непрямым движением, что так же создает аварийную ситуацию из-за недостаточного угла обзора у водителей.

Рассмотрим участок от улицы Туполева до пл. Громова (бывшей пл. Кирова) на данном участке дороги нейросеть отметила 2 места с повышенной опасностью что связано с тем что большой поток машин из 4 полос перестраивается в 1 полосу. Так же улица плохо освещена, о чем свидетельствует статистика ДТП 2 наезда и 1 авария произошли до 9:00 утра. А на площади Громова круговое движение затрудняется тем, что отсутствует регулировка движения знаками приоритета и светофорами, большой поток машин с 2 главных улиц города и Рязанского Шоссе сильно загружает единственную полосу на пути в соседний город Раменское. Для решения данной проблемы властями было принято решение о строительстве дополнительной развязки.

Исходя из проведенных исследований можно сделать несколько предложений по улучшению дорожной ситуации в городе Жуковский:

- Увеличить количество лежащих полицейских для контроля скоростного режима на территории города.
- Увеличить количество и качество дорожного освещения.
- Пересмотреть расстановку дорожных знаков на участках дороги с повышенной аварийностью.
- Организовать проведение лекций, направленных на информирование горожан о правилах дорожного движения.

Подводя итоги, можно сделать вывод о том, что использование больших данных как инструмента для решения проблем в области дорожно-транспортных происшествий, может уменьшить ущерб от аварий, а также спасти множество жизней путем выявления закономерностей лежащих в основе ДТП.

## **Библиографический список**

1. Кузнецова А.И. «Инфраструктура как необходимое условие устойчивого развития инновационной экономики города» // Вестник Московского университета имени С.Ю. Витте. Серия 1: Экономика и управление, – №1 (1), – 2012, – с. 45-50.
2. Суворов С.В., Осипов А.В., Жилиева И.А. Концепция интеллектуальной транспортной системы для автономных транспортных средств на основе многоагентного подхода. // Российский экономический интернет – журнал. – 2019. – №4. – С. 1-11. [Электронный ресурс] – URL: ([http://www.e-rej.ru/Articles/2019/Suvorov\\_Osipov\\_Zhilyaeva.pdf](http://www.e-rej.ru/Articles/2019/Suvorov_Osipov_Zhilyaeva.pdf)).
3. Поляков Д.Н. Понятие и причины дорожно-транспортных происшествий. // Сибирское юридическое обозрение, – (16), 33-35, – 2011.
4. Куприяновский В.П., Суконников Г.В., Синягов С.А., Намиот Д.Е., Карасев О.И., Бубнов П.М. Цифровая трансформация экономики, железных дорог и умных городов. Планы и опыт Великобритании. // International Journal of Open Information Technologies, – 4 (10), – 22-31, 2016.
5. Показатели состояния безопасности дорожного движения [Электронный ресурс] - URL: <http://stat.gibdd.ru/> .
6. Moosavi, S, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. «A Countrywide Traffic Accident Dataset», arXiv preprint arXiv:1906.05409, 2019.
7. Moosavi S., Mohammad H. S., Srinivasan P, Radu T, and Rajiv R. «Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights». In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.
8. Савиных В.П., Цветков В.Я. Геоданные как системный информационный ресурс // Вестник Российской академии наук. – 2014. – Т. 84. – №9. – С. 826-829.
9. Cohen W.W., Richman J. Learning to match and cluster large high – dimensional data sets for data integration // Proceedings of the eighth ACM SIGKDD

international conference on Knowledge discovery and data meaning. – ACN 2002. P. 475-480.

10. IBM Big Data Platform – Bringing big data to the Enterprise. [Электронный ресурс] – URL: <https://www.ibm.com/software/data/bigdata>.

## Reference

1. Kuznetsova A.I. «Infrastructure as a necessary condition for sustainable development of the innovative economy of the city» // Bulletin of the Moscow University named after S. Yu.Witte. Series 1: Economics and management, – №1 (1), – 2012, – pp. 45-50.

2. Suvorov S.V., Osipov A.V., Zhilyaeva I.A. the concept of an intelligent transport system for Autonomous vehicles based on a multi-agent approach. // Russian economic online magazine, – 2019, – № 4, – pp. 1-11. [Electronic resource] – url: ([http://www.e-rej.ru/Articles/2019/Suvorov\\_Osipov\\_Zhilyaeva.pdf](http://www.e-rej.ru/Articles/2019/Suvorov_Osipov_Zhilyaeva.pdf)).

3. Polyakov D.N. the Concept and causes of road accidents. // Siberian legal review, – (16), 33-35, – 2011.

4. Kuprianovsky V.P., Sukonnikov G.V., Sinyagov S.A., Namiot D.E., Karasev O.I., Bubnov P.M. Digital transformation of the economy, Railways and smart cities. UK plans and experience. // International journal of open information technologies, – 4 (10), – 22-31, 2016.

5. Indicators of road safety status [electronic resource] - url: <http://stat.gibdd.ru/>.

6. Mousavi, S., Mohammad Hossein Samawatie, Srinivasan Parthasarathi and Rajiv Ramnath. «Nationwide road accident data set», Preprint arXiv arXiv:1906.05409, 2019.

7. Moosavi S., Mohammad H.S., Srinivasan P, Radu T, and Rajiv R. «predicting accident risk based on heterogeneous sparse data: a new data set and ideas». In proceedings of the 27th ACM SIGSPATIAL International conference on advances in geographic information systems, ACM, 2019.

8. Savinykh V.P., Tsvetkov V.Ya. GEODATA as a system information resource // Bulletin of the Russian Academy of Sciences, – 2014, – Vol. 84, – № 9, pp. 826-829.

9. U.W. Cohen, letter J. training to fit and cluster large multidimensional data arrays for data integration // Proceedings of the eighth international conference, ACM Symposium on knowledge discovery and data meaning. – ACN 2002. – Pp. 475-480.

10. IBM Big Data Platform – transferring big data to the enterprise. [Electronic resource] – url: <https://www.ibm.com/software/data/bigdata>.