

УДК 372.881.1

DOI: 10.46548/21vek-2021-1054-0004

АЛГОРИТМ КЛАСТЕРИЗАЦИИ И АНАЛИЗА МЕСТ ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ

© 2021

Костин Алексей Владимирович, начальник

*Управление цифрового развития, информационных технологий и связи Пензенской области
(440000, Россия, г. Пенза, ул. Московская, 75, e-mail: kostin@obl.penza.net)*

Аннотация. Одной из проблем современного общества в России и за рубежом является необходимость сокращения общего количества дорожно-транспортных происшествий (ДТП) на автомобильных дорогах и, в первую очередь, количества погибающих и получающих ранения. Этой проблеме уделяется серьезное внимание во всех развитых странах, о чем свидетельствуют многочисленные публикации и принимаемые государственные программы, направленные на её решение. Принятые в РФ на государственном уровне документы определяют порядок учета ДТП и выявления мест их концентрации, на основе которых созданы информационные ресурсы, предоставляющие достаточно детальную статистическую информацию о ДТП и значениях показателей, характеризующих их последствия. Однако разработанные Положения не в полной мере учитывают возможности современных информационных технологий в плане выявления и динамики изменений мест концентрации ДТП, а также формирования визуальных представлений для анализа влияния реализуемых мероприятий на показатели, характеризующие ДТП. Предлагается алгоритм выявления и ранжирования мест концентрации ДТП с возможностью варьирования состава показателей и факторов, влияющих на ДТП, для выполнения разностороннего анализа мест и причин ДТП и формирования данных для выработки решений по планированию мероприятий, направленных на улучшение состояния дел в области организации безопасности дорожного движения (ОБДД). Предлагаемый алгоритм основан на алгоритме кластеризации *FOREL*, позволяет формировать наборы данных о местах концентрации ДТП в формате СУБД *PostgreSQL*, которые используются в геоинформационном портале ситуационного центра для представления на электронных картах, что сокращает продолжительность процесса анализа результатов ДТП и выработки решений по повышению уровня безопасности дорожного движения.

Ключевые слова: дорожно-транспортные происшествия, ДТП, места концентрации ДТП, алгоритм анализа мест ДТП, кластеризация, информационные ресурсы.

ALGORITHM FOR CLUSTERING AND ANALYZING TRAFFIC ACCIDENT LOCATIONS

© 2021

Kostin Aleksey Vladimirovich, head

*Administration of digital development, information technologies and communications of the Penza region
(440000, Russia, Penza, Moskovskaya St., 75, e-mail: kostin@obl.penza.net)*

Abstract. One of the problems of modern society in Russia and abroad is the need to reduce the total number of road accidents on the roads and, first of all, the number of people killed and injured. This problem is given serious attention in all developed countries, as evidenced by numerous publications and government programs aimed at solving it. The documents adopted in the Russian Federation at the state level determine the procedure for accounting for road accidents and identifying places of their concentration, on the basis of which information resources are created that provide sufficiently detailed statistical information about road accidents and the values of indicators that characterize their consequences. However, the developed Provisions do not fully take into account the capabilities of modern information technologies in terms of identifying and dynamics of changes in the places of concentration of road accidents, as well as the formation of visual representations for analyzing the impact of implemented measures on the indicators that characterize road accidents. An algorithm is proposed for identifying and ranking the places of concentration of road accidents with the possibility of varying the composition of indicators and factors affecting road accidents, for performing a comprehensive analysis of the places and causes of road accidents and generating data for developing decisions on planning measures aimed at improving the state of affairs in the field of road safety management. The proposed algorithm is based on the *FOREL* clustering algorithm, which allows you to generate data sets about the places of concentration of accidents in the *PostgreSQL* DBMS format, which are used in the geographic information portal of the situation center for presentation on electronic maps, which reduces the duration of the process of analyzing the results of accidents and developing solutions to improve road safety.

Keywords: traffic accidents, places of concentration of accidents, algorithm for analyzing accident sites, clusterization, information resources.

Введение. Статистика по количеству ДТП и их последствиям указывает на необходимость проведения систематической работы по обеспечению безопасности дорожного движения (БДД), включая мониторинг

данных о ДТП, планирование и реализацию соответствующих мероприятий. Об этом свидетельствуют данные о ДТП, в частности:

– по данным Всемирной организации здравоохра-

нения (ВОЗ) ежегодно в результате дорожно-транспортных происшествий во всем мире погибает около 1,4 миллиона человек и 45 миллионов получают травмы. По оценке ВОЗ (2013 года) в 2020 году на дорогах по всему миру могли погибнуть 17 миллионов человек и более 200 миллионов человек могли получить серьезные и необратимые травмы [1, 2];

– за последние три года (в 2018 – 2020 гг.) количество погибших в ДТП в РФ составляет 18214, 16981 и 16152 человек, соответственно, а раненых – 214853, 210877 и 183040 человек (при наличии тенденций к сокращению количества погибших и раненых) [3];

– за последние три года (в 2018 – 2020 гг.) количество погибших в ДТП в Пензенской области составляет 212, 204 и 213 человек, соответственно (при отсутствии тенденции к сокращению количества погибших), а раненых – 2560, 2525 и 2179 человек (при наличии тенденции к сокращению количества раненых) [3].

Поэтому использование информационных технологий для решения задач учета ДТП, анализа ДТП и их последствий, выявления мест концентрации ДТП, анализа эффективности мероприятий, направленных на обеспечение БДД, является актуальным, т.к. позволяет оперативно формировать данные и визуальные представления с использованием электронных карт и диаграмм, выполнять многомерный анализ данных для выявления зависимостей ДТП от различных факторов, формировать данные для выявления тенденций в сфере ОБДД.

Поскольку вопросами ОБДД в РФ занимаются на всех уровнях системы управления государством, то был выполнен анализ соответствующих нормативных документов обязательного и рекомендательного характера, в которых определен порядок учета ДТП и взаимодействия участников процесса ОБДД [4, 5], направлений работ по совершенствованию ОБДД [6], рекомендации по выполнению работ в части повышения качества автомобильных дорог и их обустройства [7, 8]. Для регистрации данных о ДТП, формирования соответствующих статистических данных по регионам, созданы информационные ресурсы, предоставляющие достаточно детальную статистическую информацию о ДТП и значениях показателей, характеризующих их последствия [3, 9].

Однако указанные документы не в полной мере учитывают возможности современных информационных технологий в плане выявления мест концентрации ДТП с учетом различных показателей и их комбинаций, а также динамики изменений (миграции) мест концентрации ДТП, формирования визуальных представлений для анализа влияния реализуемых мероприятий на показатели, характеризующие ДТП. В частности, описанные в [7, 8] методы выявления мест концентрации ДТП, ориентированные на неавтоматизированную обработку данных, сложно применять из-за значительной трудоемкости, и их можно считать морально устаревшими.

Важное значение для создания условий безопас-

ного дорожного движения имеет реализация информационного взаимодействия соответствующих организаций: органов исполнительной власти и местного самоуправления, органов МВД, медицинских учреждений, предприятий по строительству и обслуживанию дорог, средств массовой информации, диспетчерских служб, служб МЧС, общественных организаций [5, 6]. Опыт других стран показывает различные подходы к решению этих задач, например: использование математической модели, включающей различные методы идентификации «горячих точек» для программы безопасности на дорогах Турции при наличии данных не менее чем 7 лет [2]; прогнозирование продолжительности устранения инцидентов, произошедших на автострате с учетом существенных факторов, и организация оперативного информационного взаимодействия участников дорожного движения [10, 11]; использование моделей прогнозирования ДТП на пешеходных переходах [12].

Для выявления мест концентрации ДТП можно использовать подходы, которые базируются на моделях и алгоритмах кластеризации объектов, как универсальных [13 – 15], так и специализированных [16 – 19]. Универсальные алгоритмы кластеризации выполняют анализ всех объектов, что не требуется в случае выявления мест территориальной концентрации ДТП, т.к. в первую очередь надо определить места, в которых произошло не менее трёх ДТП, а одиночные случаи рассматриваются только в случаях, если эти места находятся «в прямой причинно-следственной связи с недостатками транспортно-эксплуатационного состояния» [7]. Поэтому для разработки программного обеспечения, предназначенного для выявления мест концентрации ДТП, необходима разработка специализированного алгоритма оригинального или на базе известных. В качестве прототипа выбран алгоритм *FOREL* [14].

Материалы и результаты исследования. Задачу определения мест концентрации ДТП можно рассматривать как задачу кластеризации объектов, имеющих географические координаты, т.е. кластер должен включать наиболее близкие места ДТП с учетом ограничений на размеры таких мест: не более $L_{дон}^A = 1$ км на автодорогах вне населенных пунктов и не более $L_{дон}^{НП} = 200$ м в населенных пунктах [5, 7, 8]. Соответственно, координаты ДТП, отнесенных к i -му кластеру, должны быть от центра кластера не далее, чем половина $L_{дон}^A$ и $L_{дон}^{НП}$.

Для решения задачи исходные данные должны содержать следующие обязательные данные: идентификационный номер ДТП, дата и время события, географические координаты места ДТП, показатели последствий ДТП (количество погибших и раненых). Другие классификационные признаки (факторы, которые могут влиять на ДТП) в данной постановке не требуются. Принято, что необходимые классификационные признаки использованы при отборе данных для решения данной задачи.

Кластеризация мест ДТП должна обеспечивать

выявление мест концентрации ДТП (очагов аварийности, ОА) по данным, полученным в течение заданного интервала времени (Дата1, Дата2) по региону в целом или в пределах заданной территории, которая может быть задана на карте в форме многоугольника.

Для определения ОА следует задать «радиус ОА» (R) – расстояние от центра ОА до наиболее удаленного места ДТП, включенного в данный ОА: R не более L_{don} , A или $L_{don, НГ}$

Очаги аварийности принято оценивать по значениям единичных и комплексных показателей:

- количество погибших (N_n);
- количество раненых (N_p);
- общее количество ДТП (N_o);
- «степень тяжести последствий»: $S_{mn} = 100 N_n / N_p$ [7];
- обобщенный (комплексный) показатель аварийности i -го ОА:

$Q_i = \sum_{j=1}^3 k_j w_{ij}^*$, где k_1, k_2, k_3 – весовые коэффициенты нормированных частных показателей очагов аварийности (w_{ij}^*):

$$w_{ij}^* = \frac{p_{ij} - p_{j, \min}}{p_{j, \max} - p_{j, \min}};$$

p_{ij} – j -й частный показатель для i -го кластера;

$p_{ij}^{\min}, p_{ij}^{\max}$ – наименьшее и наибольшее значения j -го частного показателя (по всем сформированным кластерам). В составе Q_i учитываются частные показатели N_n, N_p, N_o . При этом должно выполняться условие: $\sum_{j=1}^3 k_j = 1$.

Для анализа очагов аварийности следует использовать различные варианты значений весовых коэффициентов (табл. 1):

- если требуется сравнить ОА по количеству погибших в ДТП, то следует выбрать первый вариант значений коэффициентов;
- если требуется сравнить ОА по количеству пострадавших (раненых), то – второй;
- если – без учета погибших и раненых, то – третий;
- если требуется сравнить ОА по количеству погибших и раненых, то – четвёртый (при этом пользователь может задать равные предпочтения частных показателей при $k_1 = k_2 = 0,5$ или разные, например, $k_1 = 0,7, k_2 = 0,3$;
- если требуется сравнить ОА с учетом трёх показателей, то – пятый (при этом пользователь должен задать предпочтения частных показателей, например, $k_1 = 0,6, k_2 = 0,3, k_3 = 0,1$).

Таблица 1 – Варианты назначения весовых коэффициентов

Вариант	Частные показатели, учитываемые в составе Q	k_1	k_2	k_3
1	Количество погибших	1	0	0
2	Количество раненых	0	1	0
3	Количество ДТП без погибших и раненых	0	0	1
4	Количество погибших и раненых	>0	>0	0
5	Количество ДТП, количество погибших и раненых	>0	>0	>0

Алгоритм выявления мест концентрации ДТП. Алгоритм кластеризации ДТП включает следующие процедуры:

1. Формирование из исходных данных о ДТП выборки данных (N записей) для анализа ДТП в заданном интервале времени и на заданной территории.

2. Вычисление расстояний между всеми местами ДТП; результат можно представить в матрице $L_{N \times N}$ или в форме списка расстояний между ДТП: номер ДТП (i), номер ДТП (j), расстояние между i -м и j -м ДТП (L_{ij}). Можно для работы сформировать список расстояний, из которого удалены все строки с расстояниями, превышающими величину R .

3. Поскольку в начале работы алгоритма кластеры ещё не сформированы, то выбор выполнить поиск элемента с наименьшим значением ($L_{\min} = L_{ij}$) во всей матрице (или во всем списке), т. е. найти пару наиболее близко расположенных друг к другу мест ДТП.

Если $L_{\min} > R$, то вывести сообщение: «Нет ОА с радиусом не более R ; далее следует задать новое значение R , превышающее предыдущее, и вернуться к п. 3 или закончить работу.

Если найдено $L_{\min} \leq R$, то номер кластера $C=0$, перейти к п. 4.

Создать запись с данными первого/очередного кластера (табл. 2). Состав кластера можно формировать в форме отдельной таблицы.

4. При создании новой записи для кластера в таблицу записать:

- идентификатор кластера/ОА: $C=C+1$;
- номер строки (i) и номер столбца (j) матрицы (или идентификаторы соответствующих мест ДТП); в матрице $L_{N \times N}$ пометить столбцы (i и j), как запрещенные для выбора минимального значения ($L_{\min} = L_{ij}$) при последующем формировании текущего или будущих кластеров, т.к. соответствующие элементы уже включены в кластеры.

При использовании списка расстояний между ДТП надо удалить соответствующие строки с номерами мест, включенных в кластер.

Таблица 2 – Структура записей с данными о кластерах

Идентификатор кластера	Координаты центра кластера		Обобщенный показатель, $Q(C)$, при заданных k_1, k_2, k_3	Состав кластера (№№ ДТП или номера строк/столбцов из матрицы $L_{N \times N}$)					
	$X(Z_C)$	$Y(Z_C)$							
C_1	$X(Z_{C_1})$	$Y(Z_{C_1})$		$C_{11}=i$	$C_{12}=j$...	$C_{1,n1}$		
...
C_m	$X(Z_{C_m})$	$Y(Z_{C_m})$		$C_{m,1}$	$C_{m,2}$...	$C_{m,nm}$		

5. Вычислить координаты центра кластера:

$$X(Z_k) = \frac{1}{|\{C_k\}|} \sum_{i \in C_k} X_i;$$

$$Y(Z_k) = \frac{1}{|\{C_k\}|} \sum_{i \in C_k} Y_i.$$

6. Цикл для формирования очередного кластера:

а. Поиск в строках матрицы $L_{N \times N}$ номера которых относятся к очередному формируемому кластеру, элемента с наименьшим значением ($L_{\min} = L_{ij}$), за исклю-

чением помеченных столбцов, номера которых есть в созданных кластерах и в формируемом кластере;

б. Если $L_{\min} \leq R$, то вычислить координаты «потенциально нового» центра кластера (Z_n), иначе перейти к п. 7 для начала формирования нового кластера;

с. Вычислить расстояния от Z_n до всех точек, входящих в i -й кластер, включая «потенциально новое» место ДТП ($L(Z_{n,xy})$);

д. Если все $L(Z_{n,xy}) \leq R$, то включить «потенциально новое» место ДТП в состав i -го кластера и перейти к п. 6.а, иначе перейти к п. 7. Можно использовать следующий вариант: если для некоторого ранее включенного в формируемый кластер места выполняется условие $L(Z_{n,xy}) > R$, то можно исключить из формируемого кластера такое ранее включенное место, а включить «потенциально новое» место и выполнить п. 6.д, как указано выше; результаты кластеризации будут иными (для выбора предпочтительно варианта с точки зрения практической полезности следует выполнить статистическое исследование алгоритма).

7. Выполнить поиск элемента с наименьшим значением ($L_{\min} = L_{ij}$) в строках и столбцах матрицы расстояний $L_{N \times N}$, номеров которых нет в созданных кластерах (в непомеченных строках и столбцах).

8. Если $L_{\min} > R$, то вывести сообщение: «Сформированы m кластеров (определены m очагов аварийности с радиусом не более R)» и закончить формирование кластеров; перейти к п. 9 (вычисление значений обобщенного показателя Q для сформированных кластеров), иначе перейти к п. 4.

9. Вычисление значений обобщенного показателя Q для сформированных кластеров и среднего значения по множеству m кластеров

$$Q_{cp} = \frac{1}{m} \sum_{i=1}^m Q_i$$

10. Можно вычислить расстояния между «точками» в каждом кластере, найти наибольшее значение для каждого кластера и принять это число за диаметр круга для отображения на карте очагов аварийности. Но можно этого не делать, а обозначать OA на карте кругами с радиусом, соответствующим заданному значению R .

11. Вывести данные о кластерах ДТП:

– в форме таблицы, в которой указать количество мест ДТП в каждом OA вместо перечисления мест; выделить кластеры, у которых $Q_i \geq Q_{cp}$ (цветом строки или символом «>» в отдельном столбце);

– в картографической форме: места OA (координаты центров всех сформированных кластеров) или места OA , у которых $Q_i \geq Q_{cp}$; OA обозначить кругами с радиусом R или с «фактическим» диаметром (см. п. 10).

12. Вывести общие сведения о результатах формирования кластеров:

– координаты центров кластеров, которые соответствуют центрам мест концентрации ДТП, формируемых в соответствии с задаваемыми параметрами и показателями (радиус кластера, показатель ДТП) с выделением кластеров, значение показателя которых

превышает заданное значение;

– общее количество мест ДТП, сгруппированных в кластеры (α) по сделанной выборке мест ДТП;

– отношение $(100\alpha/N)\%$.

Блок-схема алгоритма соответствует описанию, начиная с п. 2. Анализ структуры алгоритма показывает, что временная сложность составляет не более $O(N^2)$, что соответствует достаточно часто применяемым для решения задач кластеризации программным средствам [16].

Заключение. На основе поставленной задачи выявления мест концентрации ДТП сформулирована задача кластеризации заданных объектов с учетом заданных факторов и показателей, разработана структура данных и разработан алгоритм кластеризации заданных объектов, позволяющий в автоматическом режиме формировать данные для последующего анализа мест ДТП с использованием электронных карт и других визуальных представлений. Выполнена оценка временной сложности разработанного алгоритма.

На основе предложенного алгоритма разработан программный модуль, который находится на стадии опытной эксплуатации и позволяет формировать наборы данных о местах концентрации ДТП в формате СУБД *PostgreSQL* для заданных территорий региона и заданных интервалов времени. Эти наборы данных, как и первичные данные о ДТП, используется в геоинформационном портале ситуационного центра Губернатора Пензенской области для представления мест концентрации ДТП на электронной карте региона для анализа и выработки решений по планированию мероприятий, направленных на повышение безопасности дорожного движения в регионе.

СПИСОК ЛИТЕРАТУРЫ:

1. Повышение безопасности дорожного движения во всем мире. Доклад Всемирной организации здравоохранения на Генеральной Ассамблее ООН 24.08.2017 г. – 25 с. [Электронный ресурс] – URL: https://unece.org/fileadmin/DAM/trans/doc/2017/wp1/UNSG_Report_72-359_ru.pdf (дата обращения 11.03.2021).
2. Saffet Erdogan, Veli İlçi, Omer M. Soysal, Aysegul Kormaz. A model suggestion for the determination of the traffic accident hotspots on the turkish highway road network: a pilot study. [Электронный ресурс] – URL: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1982-21702015000100169&lng=en&tlng=en (дата обращения 04.11.2020).
3. Показатели безопасности дорожного движения. [Электронный ресурс] – URL: <http://stat.gibdd.ru/> (дата обращения 11.03.2021).
4. Постановление Правительства РФ от 19 сентября 2020 г. № 1502 «Об утверждении Правил учета дорожно-транспортных происшествий, об изменении и признании утратившими силу некоторых актов Правительства Российской Федерации».
5. Приказ МВД РФ от 19 июня 2015 г. № 699 «Об организации учета, сбора и анализа сведений о дорожно-транспортных происшествиях».
6. Паспорт национального проекта «Безопасные и качественные автомобильные дороги» (утвержден Президиумом Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам 24 декабря 2018 года).
7. ОДМ 218.6.015–2015. Рекомендации по учету и анализу дорожно-транспортных происшествий на автомобильных дорогах Российской Федерации. – М.: Федеральное дорожное агентство (Росавтодор), 2015. 82 с.

8. ОДМ 218.4.004-2009. Руководство по устранению и профилактике возникновения участков концентрации ДТП при эксплуатации автомобильных дорог. – М.: Федеральное дорожное агентство (Росавтодор), 2009. – 94 с.
9. АИУС ГИБДД – оперативная помощь при ДТП. [Электронный ресурс] – URL: <https://dorpeh.ru/avtomobilnoe-pravo/gibdd/aius-gibdd-operativnaya-pomoshh-pri-dtp> (дата обращения 11.03.2021).
10. Traffic Incident Management Handbook. Report № HWA-HOP-10-013. - Science Applications International Corporation (SAIC), American Transportation Research Institute, 2010. – 116 pp.
11. Jinjun Tang, Lanlan Zheng, Chunyang Han, Fang Liu, Jianming Cai. Traffic Incident Clearance Time Prediction and Influencing Factor Analysis Using Extreme Gradient Boosting Model // Hindawi. Journal of Advanced Transportation, Volume 2020, Article ID 6401082, 12 pages. [Электронный ресурс] – URL: <https://doi.org/10.1155/2020/6401082> (дата обращения 11.03.2021).
12. Piotr Olszewski, Beata Osńska, Piotr Szagała, Paweł Włodarek. Development of accident prediction models for pedestrian crossings // MATEC Web of Conferences 231, 03002 (2018) [Электронный ресурс] – URL: <https://doi.org/10.1051/mateconf/201823103002> ().
13. Буреева Н. Н. Многомерный статистический анализ с использованием ППП “STATISTICA”. Учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». Нижний Новгород, 2007. – 112 с.
14. Лекции по алгоритмам кластеризации и многомерного шкалирования. К. В. Воронцов. 21 декабря 2007 г. – [Электронный ресурс] URL: <http://www.ccas.ru/voron/download/Clustering.pdf> (дата обращения 15.02.2021).
15. Тюрин А.Г., Зуев И.О. Кластерный анализ, методы и алгоритмы кластеризации // Вестник МГТУ МИРЭА, 2014, № 2, июнь 2014, выпуск 3, С. 86 – 97.
16. DBSCAN. Основанная на плотности пространственная кластеризация для приложений с шумами [Электронный ресурс] – URL: <https://ru.wikipedia.org/wiki/DBSCAN> (дата обращения 15.11.2020).
17. Li, Ruimin; Pereira, Francisco Camara; Ben-Akiva, Moshe E (2018). Overview of traffic incident duration analysis and prediction. European Transport Research Review, 2018, pp. 10 - 22. . [Электронный ресурс] – URL: <https://etr.springeropen.com/articles/10.1186/s12544-018-0300-1> (дата обращения 04.03.2021).
18. Meraldo A. Live Prediction of Traffic Accident Risks Using Machine Learning and Google Maps. [Электронный ресурс] – URL: <https://towardsdatascience.com/live-prediction-of-traffic-accident-risks-using-machine-learning-and-google-maps-d2ceffb9389e> (дата обращения 24.02.2021)
19. Д. С. Шалымов. Алгоритмы устойчивой кластеризации на основе индексных функций и функций устойчивости. - [Электронный ресурс] URL: <https://math.spbu.ru/user/gran/soi4/shalymov4.pdf> (дата обращения 01.03.2021).

Статья поступила в редакцию 11.03.2021

Статья принята к публикации 16.06.2021