

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**  
**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра информационных технологий**

**КУРСОВАЯ РАБОТА**

**МАТЕМАТИЧЕСКАЯ МОДЕЛЬ КАУЗАЛЬНОГО АНАЛИЗА  
ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ**

Работу выполнил \_\_\_\_\_ Д.Х. Чан  
(подпись)

Направлению подготовки 01.03.02 – «Прикладная математика и информатика»  
курс 3

Направленность (профиль) «Системное программирование и компьютерные  
технологии» (Математическое и программное обеспечение вычислительных  
машин)

Научный руководитель  
канд. техн. наук, доц. \_\_\_\_\_ А.Н. Полетайкин  
(подпись, дата)

Нормоконтролер ст. преп. \_\_\_\_\_ А.В. Харченко  
(подпись, дата)

Краснодар

2019

## **РЕФЕРАТ**

Курсовая работа 32 с., 9 рис., 13 источников.

**ДСМ – МЕТОД АПГ, ДТП, АВАРИЙНАЯ СИТУАЦИЯ, ПРИЧИННО-СЛЕДСТВЕННАЯ СВЯЗЬ, БЕЗОПАСНОСТЬ ДОРОЖНОГО ДВИЖЕНИЯ**

Цель работы: установить причинно-следственную связь ДТП и разработать программу для прогнозирования аварийных ситуаций и повышения безопасности дорожного движения в городе в будущем.

## СОДЕРЖАНИЕ

	4
1 Постановка задачи.....	6
2 Обзор существующих работ в решении задач рассматриваемого типа.....	7
2.1 ДСМ–метод .....	7
2.1.1 Архитектура ДСМ-системы.....	14
2.2 Суть ДСМ-метода.....	15
2.3 Объектная модель причинности.....	17
2.4 Применение ДСМ-метода для решения этой задачи.....	19
3 Описание используемых данных.....	23
3.1 Исходные данные.....	23
3.2 Конечный результат.....	25
4 Структура программы.....	27
4.1 Разработка программы.....	27
4.2 Результаты исследования.....	29
Заключение.....	30
Список использованных источников.....	31

## ВВЕДЕНИЕ

Уже на протяжении нескольких лет все серьезнее встает проблема большого количества автотранспорта в городе и, как следствие, роста дорожно-транспортных происшествий. Среди всех регионов Российской Федерации наиболее высокая смертность от дорожно-транспортных происшествий зафиксирована в Краснодарском крае. В рейтинге городов с самой высокой смертностью Краснодар занимает первое место. В столице края показатель смертельных случаев при авариях составил 7,1 на 100 тыс. населения. По данным с начала года в крае зафиксировали 222 ДТП [1].

Целью курсовой работы является исследование отношений зависимости в системе переменных, которое заключается в проверке соответствия выявленных причинных связей определенным данным для повышения безопасности движения на дорогах города. Одной из широко известных таких методик для выявления и пристального изучения причинно-следственных зависимостей и определения закономерностей возникновения ДТП является ДСМ-метод, названный в честь Джона Стюарта Миля (JSM-method), который в настоящее время широко используется в области анализа социологических данных. Этот метод поиска закономерностей по множествам положительных и отрицательных примеров опирается на методы индукции. Для решения данной задачи были использованы понятия, которые лежат в основе ДСМ-метода автоматического порождения гипотез (АПГ) и ориентированные на поиск зависимостей в эмпирических данных и использование каузальных (причинных) зависимостей для прогнозирования ДТП в дальнейшем. В этой работе предложен метод, на базе которого приводится разработка основных алгоритмов, позволяющих решить поставленную задачу, а именно алгоритмов поиска глобальных сходств в исходных данных, алгоритмов отбора гипотез из полученных сходств, а

также алгоритмов классификации новых объектов на основе полученных гипотез.

В связи с вышенаписанным, тема курсовой работы направлена на анализ аварийных ситуаций для повышения безопасности дорожного движения.

## **1 Постановка задачи**

В рамках данной курсовой работы были поставлены следующие задачи:

- обзор существующих методов анализа и изучения причинно-следственной связи между причинами и последствиями ДТП;
- изучение возможностей языка программирования C++ и его различных библиотек;
- разработка программы, позволяющая определять закономерности возникновения дорожно-транспортных происшествий.

## **2 Обзор существующих работ в решении задач рассматриваемого типа**

### **2.1 ДСМ-метод**

ДСМ-метод автоматического порождения гипотез (ДСМ-метод АПГ) был предложен В. К. Финном в конце семидесятых годов. Название метода составляют инициалы известного английского философа, логика и экономиста Джона Стюарта Милля, чьи “методы здравомыслящего естествоиспытателя” частично формализованы в ДСМ-методе.

Исторически, первым примером задач, для решения которых применялись ДСМ-системы, является выявление причинно-следственных закономерностей вида структура-активность в фармакологии.

На данный момент интеллектуальные системы типа ДСМ созданы для различных предметных областей: фармакологии, медицинской диагностики (доказательной медицины), социологии (когнитивной социологии), криминалистики, атрибуции исторических источников, робототехники, и др.

ДСМ-система предназначена для обнаружения связи между структурой объекта и его поведением. Эта связь интерпретируется как причинно-следственная связь. Как правило, характерные черты структуры объекта считаются причинами особенностей его поведения, но в некоторых случаях, принята обратная интерпретация: совокупность элементов поведения считается причиной множества характеристик его структуры.

Элементы структуры объекта будем называть атомами. Элементы поведения будем называть (целевыми) свойствами. Исследуемый объект представляется в виде конечного множества атомов. Он может обладать (или не обладать) некоторым множеством целевых свойств. Предполагается, что как у наличия, так и у отсутствия набора целевых свойств может быть причина (не обязательно единственная), эта причина является фрагментом

структуры объекта. Модель причинно-следственных связей, используемая ДСМ-методом, может быть описана графом, изображенном на рисунке 1.

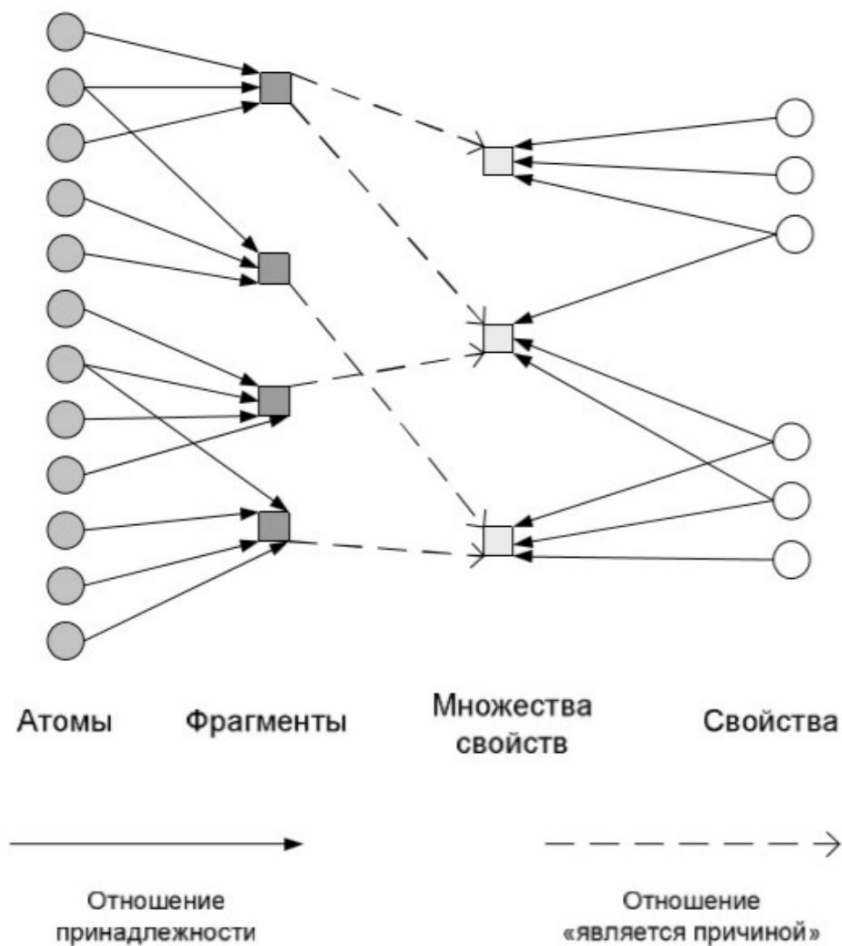


Рисунок 1 - Сущности ДСМ-метода и связи между ними

ДСМ-система решает две основные задачи:

(1) формирование гипотез о возможных причинах наличия (отсутствия) свойств у объектов (возможные причины являются фрагментами — подмножествами — объектов),

(2) формирование гипотез о наличии (отсутствии) свойств у тех объектов, для которых это было неизвестно.

Классический алгоритм ДСМ-метода, включающий работу ДСМ-системы, представляет собой итеративную процедуру, которую можно



изобразить диаграммой деятельности UML (рисунок 2). Рассмотрим основные блоки этой процедуры.

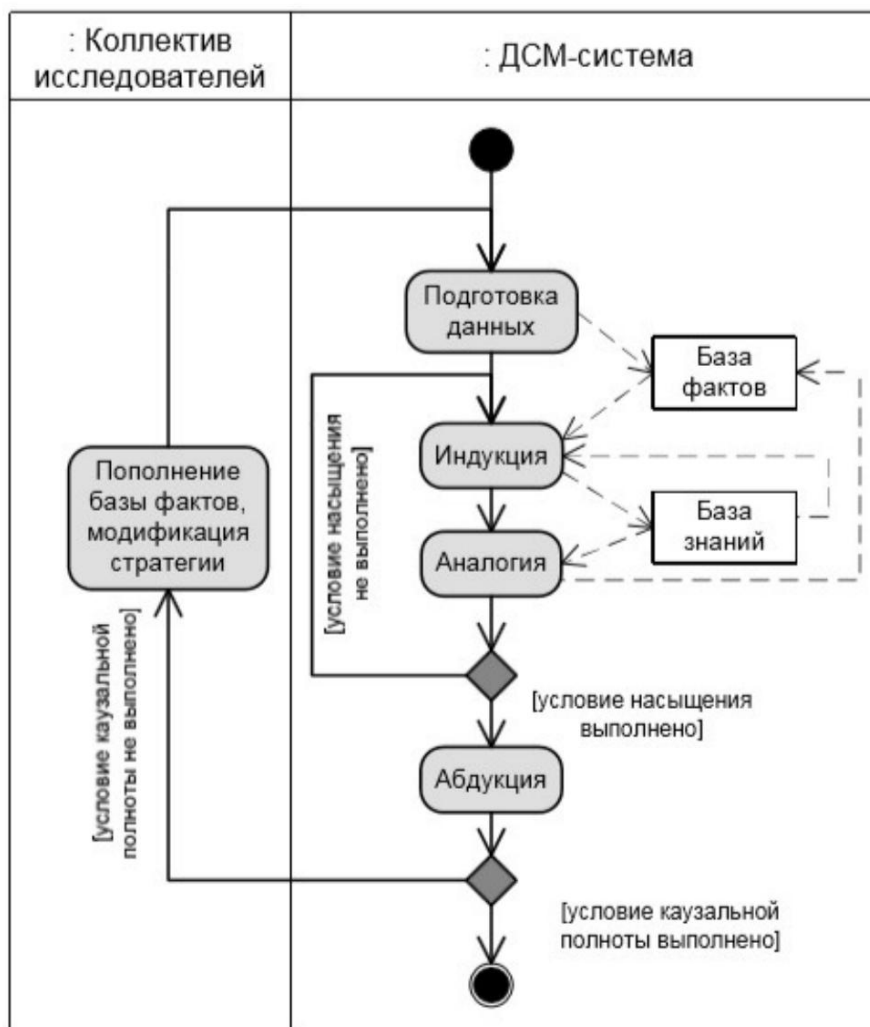


Рисунок 2 – Алгоритм ДСМ-метода

Подготовка данных в контексте работы ДСМ-системы означает преобразование данных во внутренний формат ДСМ-системы. Исходные данные могут быть представлены в некотором стандартном формате, например в виде CSV-файла, где каждому объекту соответствует строка таблицы. Для работы ДСМ-системы объекты необходимо представить в виде

множеств. В результате процедуры подготовки данных формируется база фактов, представленных во внутреннем формате ДСМ-системы.

Индукция, т.е., применение формализованных правил индуктивных рассуждений, предназначена для порождения гипотез о возможных причинах наличия (отсутствия) целевых свойств. Гипотезы о возможных причинах представляют некоторые общие закономерности, которые включаются в базу знаний ДСМ-системы.

Аналогия, т.е., применение формализованных правил рассуждений по аналогии, предназначена для порождения гипотез о наличии (отсутствии) целевых свойств у тех объектов, для которых это неизвестно. Гипотезы о наличии (отсутствии) целевых свойств имеют такой же формат, как исходные факты, и включаются в базу фактов.

Проверка условия насыщения: условие насыщения считается выполненным, если невозможно получить новые гипотезы о наличии (отсутствии) свойств у тех объектов, для которых это было неизвестно. Поскольку количество таких объектов конечно, насыщение обязательно (рано или поздно) наступает. Если насыщение наступило, происходит выход из цикла Индукция–Аналогия и переход к блоку абдукции.

Абдукция — это проверка условия (критерия) каузальной полноты. Условие каузальной полноты считается выполненным, если для каждого объекта, обладающего (не обладающего) требуемым набором целевых свойств найдена причина наличия (отсутствия) этого набора свойств. Если условие каузальной полноты выполнено, то работа ДСМ-системы завершается, в противном случае происходит расширение исходного набора данных и, возможно, происходит модификация стратегии работы ДСМ-системы.

Пополнение базы фактов и модификация стратегии производится пользователями и администраторами (настройщиками) ДСМ-системы. Пополняется набор исходных данных, который преобразуется в базу фактов уже самой ДСМ-системой. Модификация стратегии включает различные

операции, в частности, может быть выбрана альтернативная разновидность ДСМ-метода, изменены параметры правил индукции и аналогии и т.п.

ДСМ-рассуждением будем называть часть алгоритма ДСМ-метода, полученную удалением процедур подготовки данных и пополнения базы фактов и модификации стратегии. Эта часть относится только к ДСМ-системе и играет в ДСМ-системе очень важную роль. Введем некоторые понятия, относящиеся к работе ДСМ-рассуждению, следуя формулировкам из статьи В.К.Финна [2]:

- шагом ДСМ-рассуждения будем называть однократное выполнение процедуры индукции или процедуры аналогии.

- тактом ДСМ-рассуждения будем называть однократное последовательное выполнение процедур индукции и аналогии.

- этапом I ДСМ-рассуждения будем называть внутренний цикл работы ДСМ-метода, состоящий из последовательных тактов, выполняемых до тех пор, пока не будет выполнено условие насыщения.

- Этапом II ДСМ-рассуждения будем называть выполнение процедуры абдукции и проверку условия каузальной полноты.

ДСМ-метод автоматического порождения гипотез является не только средством машинного обучения в базах факторов (БФ) компьютерных систем, выполняющих функции анализа данных и предсказания изучаемых эффектов (они содержатся неявно в БФ), но он также представляет собой формализованную эвристику, которая является инструментом формализованной эпистемологии компьютерных систем, называемых интеллектуальными системами.

В.К. Финн в своей работе [3] выделяет у ДСМ-метод АПГ 5 компонент:

- условия применимости,
- ДСМ-рассуждения,
- представление знаний в виде открытых квазиаксиоматических теорий (КАТ),

- метатеоретические принципы и средства исследования рассуждений и предметных областей (в том числе дедуктивная имитация рассуждений, процедурная семантика и препроцессинг, результатом которого является выбор стратегий рассуждения и соответствующей им процедурной семантики),

- интеллектуальные системы типа ДСМ (ИС-ДСМ)

Основные достоинства ДСМ метода АПГ:

- будучи нестатистическим методом анализа данных, учитывает индивидуальные особенности изучаемых объектов исследования.

- способность порождать гипотезы на малых массивах данных благодаря выявленному сходству объектов, характеризующихся существенными параметрами.

- возможности работать с открытыми (пополняемыми) массивами данных, распознавая необходимость управляемого расширения БД, если таковая возникает.

- порождение полученных правил (причинных зависимостей) в явном виде, что обеспечивает интерпретацию результатов обучения.

Недостатки ДСМ метода АПГ:

- неприменимость при условии наличия противоречий в обучающих примерах.

- большая вычислительная сложность.

- необходимость корректного определения операции сходства на примерах.

ДСМ метод АПГ активно применяется при решении следующих задач:

- автоматизированного построения формализации знаний,

- технической диагностики,

- исследования детерминант социального поведения,

- медицинской диагностики,

- прогнозирования путей биотрансформации,

- прогнозирования химической канцерогенности и токсичности,

- обучения мобильных роботов.

Д. С. Милль сформулировал пять своих знаменитых правил индуктивного рассуждения для решения научных задач, отличных от схемы вывода перечислительной индукции. Он назвал их, соответственно, методом схождения, методом различия, соединенным методом схождения-различия, методом остатков и методом сопутствующих изменений. Ниже приведены формулировки данных методов:

1) Метод сходств. Если два или более случая имеют общим лишь одно обстоятельство, то это обстоятельство, в котором только и согласуются все эти случаи, есть причина или следствие данного явления.

2) Метод различия. Если случай, в котором исследуемое явление наступает, и случай, в котором оно не наступает, сходны во всех обстоятельствах, кроме одного, встречающегося лишь в первом случае, то это обстоятельство, в котором только и различаются эти два случая, есть следствие, или причина, или необходимая часть причины явления.

3) Соединенный метод схождения и различия. Если два или более случая возникновения явления имеют общим лишь одно обстоятельство, и два или более случая не возникновения того же явления

4) Метод остатков. Если из явления вычтешь ту его часть, которая, как известно из прежних индукций, есть следствие некоторых определенных предыдущих обстоятельств, то остаток данного явления должен быть следствие остальных предыдущих обстоятельств.

5) Метод сопутствующих. Всякое явление, изменяющееся определенным образом всякий раз, когда некоторым определенным образом изменяется другое явление, есть либо причина, либо следствие этого явления, либо соединено с ним какой-либо причинной связью.

### 2.1.1 Архитектура ДСМ-системы

Архитектура ДСМ-системы представлена на рисунке 2. Она включает: ДСМ решатель, базу фактов (данных), базу знаний и интерфейс, разделенный на два компонента: интерфейс конечного пользователя и интерфейс для настройки системы. Конечный пользователь формулирует конкретные задачи, а настройщик системы определяет стратегию и методы подготовки данных, которые будут применяться для различных задач конечного пользователя. Роль настройщика для ДСМ-системы аналогична роли инженера по знаниям для продукционных экспертных систем.

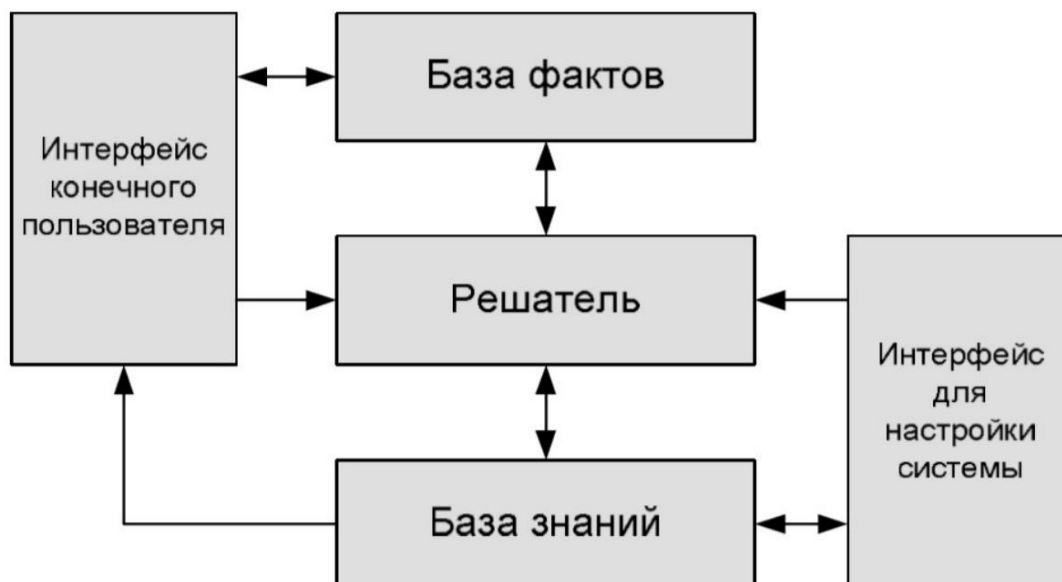


Рисунок 2 – Архитектура ДСМ – системы

База знаний ДСМ-системы имеет сложную структуру. Она включает:

- предметные знания (они описывают известные закономерности предметной области и используются, в основном, для подготовки данных), предметные знания включают правила или процедуры перевода данных во

внутренний формат и правила или процедуры для определения структуры объектов,

- метазнания — правила или процедуры, представляющие формализованные правдоподобные рассуждения, — индукцию и аналогию,
- гипотезы о причинах наличия (отсутствия) целевых свойств — они имеют формат аналогичный формату данных, но описывают общие закономерности, поэтому должны интерпретироваться как знания.

База фактов содержит:

- сведения о наличии или отсутствии целевых свойств,
- сведения о структуре объектов.

Интерфейс конечного пользователя позволяет организовать ввод исходных данных и сформулировать задание для ДСМ-системы.

Интерфейс для настройки системы позволяет модифицировать правила и/или процедуры индукции и аналогии, выбрать разновидность ДСМ-метода и задать его параметры.

Решатель исполняет ту часть алгоритма ДСМ-метода, которая относится к работе ДСМ-системы. Решатель содержит вычислитель и рассуждатель. Вычислитель исполняет процедуры подготовки данных и обнаружения сходства объектов, рассуждатель реализует процедуры правдоподобных рассуждений: индукции, аналогии и абдукции.

## 2.2 Суть ДСМ-метода

Суть ДСМ-метода состоит в следующем. Вводится три конечных множества: причин  $A = \{a_1, a_2, \dots, a_p\}$ , следствий  $B = \{b_1, b_2, \dots, b_m\}$  и оценок  $Q = \{q_1, q_2, \dots, q_l\}$ . Выражение вида  $a_i \rightarrow b_j; q_k$  называется положительной гипотезой. Оно связано с утверждением типа « $a_i$  является причиной  $b_j$ , с оценкой достоверности  $q_k$ ». Выражение вида  $a_i \rightarrow b_j; q_k$  называется отрицательной гипотезой. Оно связано с утверждением типа « $a_i$  не является причиной  $b_j$ , с оценкой достоверности  $q_k$ ». Для сокращения записи

положительные гипотезы обозначаются  $h_{ijk}^{+ii}$ , а отрицательные —  $h_{ijk}^{-ii}$ . Среди значений  $q_i$  выделяются два специальных, которые можно обозначить 0 и 1. Значение 0, приписанное положительной или отрицательной гипотезе, означает, что соответствующее утверждение является ложным. Приписывание гипотезам значения 1 означает, что данная гипотеза является тождественно истинной. Таким образом, гипотезы с оценками 0 и 1 можно рассматривать как высказывания, ложность и истинность которых твердо установлены. Все остальные оценки, отличные от 0 и 1, обозначаются рациональными числами вида  $s/n$ , где  $s$  принимает значения от 1 до  $n - 1$ . Величина  $n$  характеризует «дробность» используемых оценок достоверности.

Предположим, что при регистрации некоторого ДТП среди технических неисправностей одного из автомобилей-виновников обнаруживается неисправное рулевое управление. Через некоторое время снова в транспортном средстве, причастном к совершению ДТП, обнаруживается неисправное рулевое управление. При этом возникает положительная гипотеза вида «Если в движущемся транспортном средстве имеет место неисправное рулевое управление, то оно потенциально является виновником ДТП». Оценка достоверности этой гипотезы при двух наблюдениях будет невелика. Однако если при дальнейшем изучении этого вопроса окажется, что автомобили, причастные к совершению ДТП, имели неисправное рулевое управление, то оценка выдвинутой выше гипотезы будет все время возрастать, и все же эта оценка никогда не станет равной 1. В связи с этим введем два типа истинности: эмпирическую истину и теоретическую истину. В данном примере высказыванию о потенциальной причастности к ДТП присваивается значение эмпирической истины, так как все наблюдения были в пользу данной гипотезы. Однако вполне можно допустить, что есть небольшое количество транспортных средств с такой же неисправностью, которые не спровоцировали ДТП, и в связи с этим не попали в сферу внимания исследователя. Совсем другое положение будет в том случае, когда в правилах дорожного движения будет сказано, что



«запрещается дальнейшее движение транспортных средств, в которых рулевое управление не дает возможности водителю осуществить маневр во время движения с минимальной скоростью» [6]. При такой информации высказывание о причастности к ДТП автомобиля, движущегося с неисправным рулевым управлением, будет оценено как теоретическая истина.

Рассмотренный пример иллюстрирует процесс оценивания степени достоверности гипотезы, когда предполагаемая причина (в данном случае – неисправность автомобиля) уже выделена из множества возможных причин. В ДСМ-методе формализован не только этот этап, но и предшествующий ему этап нахождения кандидата в причины, которая могла бы вызвать интересующее следствие (в данном случае – причастность к совершению ДТП). В приведенном выше примере это соответствовало бы следующему. Наблюдая на улицах города потоки автотранспорта и выделяя в них автомобили с признаками неисправностей рулевого управления, надо «сообразить», что такие автомобили могут спровоцировать ДТП.

### 2.3 Объектная модель причинности

Причины могут быть различными по типу. Рисунок 3 демонстрирует различные типы причин и их отношения наследования относительно суперкласса *Причины*. Наиболее редкими являются *необходимые* и *достаточные причины*. Если  $a_i$  – причина такого типа, то  $b_j$  происходит всегда, и если  $b_j$  произошло, то наверняка было  $a_i$ . Примерами такой «жесткой» связи двух явлений может служить падение тела, если для него отсутствует опора. Чаще встречаются достаточные причины, всегда вызывающие появление  $b_j$ . Но появление  $b_j$  не служит стопроцентным обоснованием того, что до этого было  $a_i$ . Следствие  $b_j$  могло быть вызвано и какими-то другими достаточными причинами. Так, например, возможной причиной возникновения ДТП может быть возгорание транспортного

средства во время движения, ибо этого вполне достаточно, чтобы данная ситуация попала под определение ДТП, которое дается в п.1.2 ПДД Российской Федерации, но весьма вероятно, что могут быть какие-либо другие причины возникновения этого ДТП.



Рисунок 3 - Отношения между типами причин в нотации UML

*Дополнительные* причины обладают тем свойством, что их наличие не вызывает следствия  $b_j$ . Для того чтобы  $b_j$  появилось, нужен вполне определенный набор дополнительных причин, который выступает в роли обобщенной достаточной причины появления  $b_j$ . Легко себе представить такой набор причин, который приводит, например, к попаданию автомобиля в ДТП при движении по городу в час пик в условиях плохой видимости. Перечисление и обсуждение дополнительных причин, приведших к ДТП, – обыкновенное занятие для участников или свидетелей такого происшествия. Среди дополнительных причин могут быть *необходимые дополнительные* причины. Их вхождение в набор, образующий обобщенную достаточную причину, обязательную для того, чтобы  $b_j$  реализовалось (на рис. 3 это отношение отображено значком объединения). Остальные дополнительные причины можно назвать *факультативными* или *необязательными*. В окончательный набор могут входить те или иные комбинации необходимых и необязательных причин. Так, в случае возникновения ДТП типа «Наезд на

препятствие» две дополнительные причины являются заведомо необходимыми: 1) движение транспортного средства и 2) пересечение траектории его движения с препятствием. Остальные дополнительные причины являются факультативными. Наконец, *возможные* причины  $a_i$  обладают тем свойством, что появление  $a_i$  необязательно вызывает  $b_j$ , но увеличивает возможность появления  $b_j$ .

Кроме причин  $a_i$  важную роль в процессах реализации причинно-следственных зависимостей играют так называемые *тормоза*. Наличие тормоза наряду с причиной, вызывающей  $b_j$  в обычных условиях, приводит к тому, что  $b_j$  не появляется. Так, вышеупомянутое ДТП типа «Наезд на препятствие» может не произойти, если водитель соблюдает крайнюю осторожность, двигаясь с минимальной скоростью.

После сказанного становится ясным, что нахождение причин-кандидатов для формируемых гипотез — далеко не тривиальная задача. В положительных и отрицательных примерах эти причины скрыты в описаниях реальных объектов, обладающих или не обладающих интересующими исследователя свойствами. Из этих описаний следует выделить кандидатов в причины, а затем убедиться, что выбор оказался не случайным.

## 2.4 Применение ДСМ-метода для решения этой задачи

В работе [5] при поиске закономерности возникновения дорожно-транспортных происшествий при помощи ДСМ-метода, авторы полагают, что создание интеллектуальных систем в области исследования дорожно-транспортной аварийности, реализующих когнитивные рассуждения, вызвано потребностью формализации качественного анализа огромного количества накопленных данных о ДТП.

Рассмотрим группу положительных примеров. Находим некоторую часть описания объектов, общую для определенной совокупности примеров из этой группы. Например, обнаруживаем в значительной части случаев ДТП

вида «Наезд на пешехода» отсутствие освещения дороги в темное время суток. Тогда есть основания считать данную обобщенную причину кандидатом в причины. Таких кандидатов может оказаться несколько. Образует матрицу  $M^{+ii}$ , в которой строки соответствуют выделенным кандидатам  $a_i$ , а столбцы – интересующим нас следствиям  $b_j$ . На пересечении строк и столбцов будем записывать оценки достоверности  $q_k$  гипотез  $h_{ijk}^{+ii}$ . Для множества отрицательных примеров аналогичным образом строится другая матрица  $M^{-ii}$ , в которой содержатся оценки достоверности отрицательных гипотез  $h_{ijk}^{-ii}$ . Кандидаты в причины в матрицах  $M^{+ii}$  и  $M^{-ii}$  могут частично совпадать, так как положительные и отрицательные примеры не образуют полной выборки из всего множества возможных примеров.

На каждом шаге работы ДСМ-метода используются новые наблюдения, пополняющие множества положительных и отрицательных примеров. Эти новые наблюдения могут либо подтверждать сформированные гипотезы  $h_{ijk}^{+ii}$  и  $h_{ijk}^{-ii}$ , либо противоречить им. В первом случае оценки достоверности соответствующих гипотез увеличиваются, а во втором – уменьшаются. Механизм изменения оценок  $q_k$  может быть различным. В ДСМ-методе он устроен следующим образом. Значение  $n$  совпадает с числом имеющихся в данный момент положительных или отрицательных примеров. Таким образом, для  $M^{+ii}$  и  $M^{-ii}$  значение  $n$  может оказаться различным. С ростом  $n$  растет дробность оценок достоверности. При этом оценка  $1/n$  играет особую роль. Она соответствует полному незнанию о достоверности гипотезы. Поэтому в начальный момент  $M^{+ii}$  и  $M^{-ii}$  заполнены лишь нулями, единицами и оценками  $1/n$ . Значения истинности и лжи могут иметь гипотезы, у которых в качестве причин даны полные описания объектов, образующих множества примеров.

Если некоторая положительная или отрицательная гипотеза  $h_{ijk}^{-ii}$  имела оценку  $k/n$ , то при появлении нового примера ( $n$  при этом заменяется на  $n + 1$ ) проверяется, подтверждает или не подтверждает новый пример эту гипотезу. При подтверждении оценка  $k/n$  заменяется на  $(k + 1) / (n + 1)$ , а при

неподтверждении новым примером ранее выдвинутой гипотезы ее оценка меняется с  $k / n$  на  $(k - 1) / (n + 1)$ . Таким образом, в процессе накопления новой информации оценки гипотез либо приближаются к 0 или 1, либо колеблются между этими граничными значениями. В первом случае гипотеза может на некотором шаге (когда будет пройден некоторый априорно заданный нижний порог достоверности) исчезнуть из  $M^{+ii}$  или  $M^{-ii}$ . Во втором случае при достижении некоторого верхнего порога достоверности гипотеза может получить оценку, отражающую эмпирическую истину, и запомниться как некий установленный факт. В третьем случае, если колебания оценок достаточно сильны, может также произойти исключение сформированной ранее гипотезы из тех, которые описаны в  $M^{+ii}$  и  $M^{-ii}$ .

Новые гипотезы формируются не только на основании выделения в примерах определенного сходства (общей части в описании). Они могут использовать и метод различия (если случай, в котором исследуемое явление наступает, и случай, в котором оно не наступает, сходны во всех обстоятельствах, кроме одного, встречающегося лишь в первом случае, то это обстоятельство, в котором одном только и разнятся эти два случая, есть следствие, или причина, или необходимая часть причины явления), также сформулированный Миллем [7] и адаптированный для задач автоматического порождения гипотез в интеллектуальных системах [8]. Различие выявляется для примеров, относящихся к группам положительных и отрицательных примеров. Найденное различие служит кандидатом для гипотез, включаемых в  $M^{+ii}$  или  $M^{-ii}$ .

Кроме выявления кандидатов в причины  $a_i$  для положительных и отрицательных гипотез в описываемом методе ищутся также тормоза, наличие которых снимает влияние  $a_i$  на появление  $b_j$ . В новых версиях метода [9] в качестве  $a_i$  выступают весьма сложные утверждения, в которых отдельные части описаний объектов могут быть связаны между собой произвольными логическими выражениями, например, следующего типа:

«Если в объекте есть  $a'$  и  $a''$  и нет  $a'''$  или в объекте есть  $a''''$ , то свойство  $b$  имеет место».

Как уже было сказано, в ДСМ-методе, кроме прямой реализации идей Милля, используются еще некоторые выводы по аналогии [8, 9]. Для этого на множестве описаний объектов вводится тем или иным способом понятие сходства. Если, например, речь идет о структурных формулах химических соединений, то мерой сходства для них могут быть совпадения самих структур при различных заполнителях позиций или, наоборот, наличие в некоторых фиксированных позициях структур одинаковых элементов. Если установлено отношение сходства, то в ДСМ-методе происходит вывод по аналогии. Он осуществляется следующим способом. Если гипотеза  $h_{ijk}$  имеет оценку  $k / n$  и такова, что причина, используемая в ней, сходна с причиной в гипотезе  $h'_{ijk}$ , имеющейся в той же матрице  $M$  и оцениваемой с точки зрения достоверности значением  $1 / n$ , то на гипотезу  $h'_{ijk}$ , переносится оценка гипотезы  $h_{ijk}$  и она получает оценку достоверности  $k / n$ . Подобная процедура в ДСМ-методе называется правилом положительной аналогии.

### 3 Описание используемых данных

#### 3.1 Исходные данные

На вход ДСМ-метод получает некоторое множество изучаемых объектов и сведения об их структуре, о наличии или отсутствии у них определенных свойств, а также, в некоторых случаях, о связи между структурой объектов и их свойств. Кроме того имеется ряд целевых признаков, каждый из которых разбивает исходное множество объектов на четыре непересекающихся подмножества:

- объекты, про которые известно, что они обладают данным признаком,
- объекты, про которые известно, что они не обладают данным признаком,
- объекты, для которых существуют аргументы как за, так и против того, что они обладают данным признаком,
- объекты, о которых неизвестно, обладают они этим признаком или нет.

Причинами ДТП являются совокупности множеств факторов, природного или техногенного характера, способных оказать прямое или косвенное влияние на возникновение ДТП. В качестве множеств дополнительных и возможных причин приняты множества, такие как:

- общие сведения (день недели, час дня, время суток),
- дорожные условия (тип и состояние покрытия, искусственные сооружения, погодные условия и пр.)
- сведения об участниках ДТП (квалификация участника, пол, возраст, водительский стаж, нарушения ПДД и др.),
- сведения о транспортном средстве (ТС) (класс и тип ТС, наличие прицепа, технические неисправности ТС),
- сведения о пострадавших (род занятий и средства пассивной безопасности).

Следствиями ДТП являются события, спровоцированные возникновением ДТП и приводящие к изменению состояния участников ДТП, транспортных средств и дорожной среды на месте происшествия. В качестве множеств следствий приняты множества, такие как:

- общие сведения (вид ДТП, количество участников и ТС),
- сведения о ТС (местонахождение ТС, внешние повреждения ТС),
- сведения о потерпевших (количество пострадавших, состояние пострадавшего, тип полученных травм).

Множества причин и следствий тем или иным образом определяют каузальную картину ДТП.

Для проведения прикладного исследования будут использоваться данные, которые извлечены из карт дорожно-транспортных происшествий с пострадавшими, зарегистрированные на территории города. Для большей наглядности будут построены аналитические графики на основе извлеченных данных. Полученные данные будут внесены в базу данных, содержащую в себе множества различных причин и следствий, которые включают полные наборы по выбранным полям карт учета ДТП.

На безопасность дорожного движения влияют множество факторов, которые можно объединить в сложную динамическую систему, включающую в себя множество элементов. Таблицы этой системы содержат в себе информацию о месте происшествия (город, район, улица), общие сведения о пострадавших, состоянии дороги, покрытия, погодных условиях, выявленных недостатках на участке дороги, а также о виде происшествия, типе дорожно-транспортного происшествия. Для удобства одна таблица содержит сведения только об одном факторе. Таблицы находятся в определенных отношениях связи друг с другом, образуя целостность системы. Благодаря этим отношениям связи можно увидеть относительно полную картину случившегося происшествия, и становится понятна возможная причина, по которой случилось ДТП, что является основной



задачей данной системы. Примерная структура полученной базы данных показана на рисунке 4.

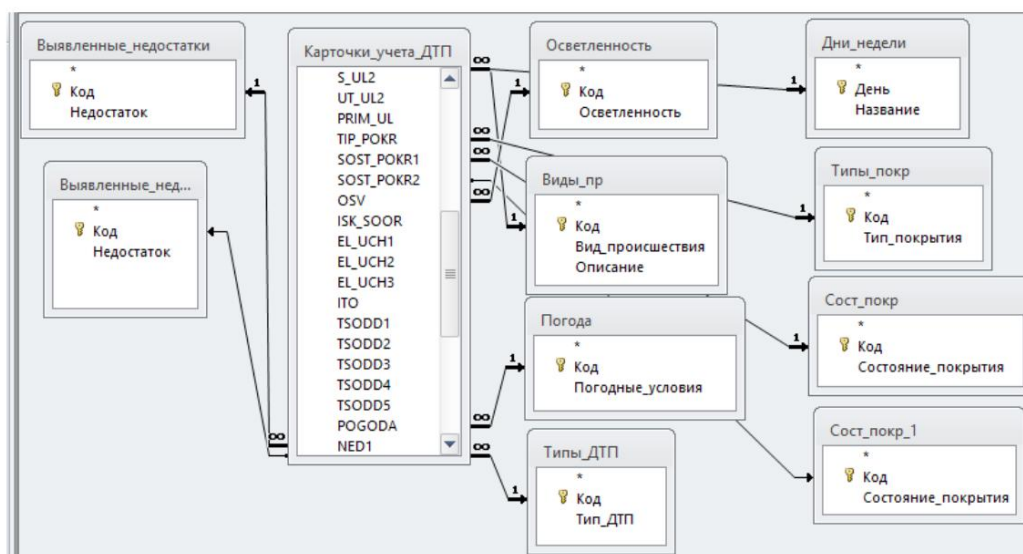


Рисунок 4 – Структура базы данных карт учета ДТП

### 3.2 Конечный результат

Результатом применения ДСМ-метода являются гипотезы двух типов:

- гипотезы о связи определенных структурных фрагментов изучаемых объектов со свойствами, которыми они обладают,
- гипотезы о наличии или отсутствии целевых признаков у объектов, для которых изначально это было неизвестно, формируемые на основании установленной взаимосвязи между свойствами объектов и их структурными компонентами.

Матрица причинно-следственных отношений – результат работы программы. На пересечении столбцов, которые являются следствиями, и строк, являющихся причинами, матриц будет оценка достоверности того является ли  $a_i$  причиной  $b_j$ . Все оценки находятся в диапазоне от 0 до 1, где значение 0 означает, что соответствующая гипотеза является ложной, а

значение 1 – гипотеза является истинной. Гипотезы с оценкой 1 или 0 рассматриваются как утверждения, истинность или ложность которых твердо установлена. В общем виде эта матрица показана на рисунке 5.

426x44	B1_01	B1_02	B1_03	B1_04	B1_05	B1_06	B1_07	B1_08	B1_09	B1_10	B2_1	B2_2	B2_3	B2_4	B3_1	B3_2	B3_3	B3_4
A1_1																		
A1_2																		
A1_3																		
A1_4																		
A1_5																		
A1_6																		
A1_7																		
A2_1																		
A2_2																		
A2_3																		
A2_4																		
A2_5																		
A2_6																		
A2_7																		
A3_1																		
A3_2																		
A3_3																		
A3_4																		
A3_5																		
A3_6																		
A3_7																		
A4_1																		
A4_2																		
A4_3																		
A4_4																		
A4_5																		
A5_1																		
A5_2																		
A5_3																		
A6_1																		
A6_2																		
A6_3																		

Рисунок 5 – Матрица причинно-следственных отношений

## 4 Структура программы

### 4.1 Разработка программы

В данном семестре было принято решение использовать Python как среду программирования и Firebird для создания и работы с базой данных.

Python — высокоуровневый язык программирования с динамической типизацией, поддерживающий объектно-ориентированный, функциональный и императивный стили программирования. Это язык общего назначения, на котором можно одинаково успешно разрабатывать системные приложения с графическим интерфейсом, утилиты командной строки, научные приложения, игры, приложения для веб и много другое [11].

Firebird - это полнофункциональная и мощная СУБД, она может обслуживать базы данных размером от нескольких килобайт до многих гигабайт, показывая хорошую производительность и практически не нуждаясь в обслуживании. Firebird включает в себя набор консольных программ, позволяющих создавать базы данных, исследовать их характеристики, выполнять операторы SQL и скрипты, производить резервное копирование данных, их восстановление из резервной копии [12].

Для начала работы требуется создать базу данных используя выгруженные данные с сайта «Показатели состояния безопасности дорожного движения» периодом в 14 дней [13]. На рисунке 6 показана структура базы.

#	PK	FK	UNQ	Field Name	Field Type	Domain	Size	s
1	1			ID	INTEGER			
2				CONSEQUENCE	VARCHAR		100	
3				WEATHER	VARCHAR		50	
4				ROADPOKRYTIE	VARCHAR		50	
5				DAYTIME	VARCHAR		50	

Рисунок 6 – Структура базы

На рисунке 7 показана часть базы данных после заполнения.

ID	CONSEQUENCE	WEATHER	ROADPOKRYTIE	DAYTIME
1	наезд на препятствие	пасмурно	сухое	темное время суток
2	наезд на препятствие	пасмурно	сухое	темное время суток
3	наезд на препятствие	пасмурно	мокрое	темное время суток
4	наезд на препятствие	пасмурно	сухое	темное время суток
5	столкновение	пасмурно	мокрое	светлое время суток
6	столкновение	дождь	мокрое	темное время суток
7	столкновение	пасмурно	мокрое	светлое время суток
8	столкновение	дождь	мокрое	темное время суток
9	наезд на пешехода	дождь	мокрое	темное время суток
10	наезд на пешехода	дождь	мокрое	темное время суток
11	наезд на пешехода	пасмурно	мокрое	темное время суток
12	наезд на пешехода	пасмурно	сухое	темное время суток
13	столкновение	пасмурно	сухое	темное время суток
14	наезд на пешехода	пасмурно	сухое	темное время суток
15	столкновение	снегопад	мокрое	светлое время суток
16	столкновение	ясно	сухое	темное время суток
17	столкновение	дождь	мокрое	темное время суток
18	наезд на пешехода	дождь	мокрое	темное время суток
19	столкновение	пасмурно	мокрое	темное время суток
20	столкновение	пасмурно	мокрое	темное время суток
21	столкновение	пасмурно	сухое	темное время суток
22	наезд на пешехода	пасмурно	сухое	темное время суток
23	опрокидывание	дождь	мокрое	темное время суток
24	столкновение	дождь	мокрое	темное время суток
25	столкновение	ясно	сухое	темное время суток

Рисунок 7 – Часть заполненной базы данных

После заполнения базы была написана программа, вычисляющая связь между различными причинами и следствиями. Код программы представлен на рисунке 8.

```
import fdb

con = fdb.connect(database='C:/Users/hicut/Downloads/ACCIDENTDB.FDB', user='SYSDBA', password='masterkey')
cur = con.cursor()

cur.execute("select ' ' from accident")
all_records = len(cur.fetchall())

cur.execute("SELECT id FROM accident WHERE consequence = 'съезд с дороги' and weather = 'гололедица'")
res = len(cur.fetchall())/all_records
print("result weather= %.3f" % res)

cur.execute("SELECT id FROM accident WHERE consequence = 'иной вид ДТП' and roadpokrytie = 'заснеженное'")
res = len(cur.fetchall())/all_records
print("result roadpokrytie= %.3f" % res)

cur.execute("SELECT id FROM accident WHERE consequence = 'наезд на стоящее ТС' and daytime = 'светлое время суток'")
res = len(cur.fetchall())/all_records
print("result daytime= %.3f" % res)
```

Рисунок 8 – Код программы

## 4.2 Результаты исследования

Для наглядности после выполнения работы программы результаты были занесены в таблицу в программе MS Excel. На рисунке 9 представлена работа программы, перенесенная в таблицу.

	пасмурно	ясно	дождь	снегопад	гололедица	сухое	мокрое	обработанное противогололедными материалами	гололедица	заснеженное	свежеуложенная поверхность	светлое время суток	темное время суток	сумерки
наезд на препятствие	0.072	0.015	0.01	0.005	0	0.026	0.078	0	0.005	0	0	0.041	0.0625	0
наезд на пешехода	0.218	0.052	0.098	0	0	0.109	0.244	0.010	0.005	0	0	0.052	0.302	0.015
наезд на велосипедиста	0.005	0	0.005	0.005	0	0	0.010	0	0	0.005	0	0	0.015	0
столкновение	0.203	0.088	0.093	0.015	0	0.114	0.265	0.010	0	0.005	0	0.182	0.213	0.005
опрокидывание	0.026	0.020	0.015	0	0	0.020	0.026	0.010	0	0.005	0	0.036	0.026	0
наезд на стоящее ТС	0.005	0	0	0	0	0	0	0	0	0.005	0	0.005	0	0
иной вид ДТП	0	0.005	0	0	0	0.005	0	0	0	0	0	0.005	0	0
съезд с дороги	0	0	0	0.005	0	0	0	0.005	0	0	0	0	0.005	0

Рисунок 9 – Результат работы программы, перенесенный в таблицу

На рисунке 9 можно увидеть, насколько сильно некоторое следствие зависит от определенной причины.

## **ЗАКЛЮЧЕНИЕ**

Благодаря ДСМ-методу АПГ возможно создание программы, которая способна анализировать причины и следствия дорожно-транспортных происшествий для улучшения безопасности на дорогах.

В рамках курсовой работы были рассмотрены основные теоретические основы изучения причинно-следственной связи ДТП с помощью ДСМ-метода и была разработана программа, которая позволяет вычислить оценку достоверности каждой возможной гипотезы.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 На Кубани с начала года в ДТП пострадали 235 детей // <https://www.kuban.kp.ru/daily/26696.7/3719840/> [4 Декабря 2018]
- 2 Финн В.К. Индуктивные методы Д.С. Милля в системах искусственного интеллекта. Часть I // Искусственный интеллект и принятие решений. — 2010. — № 3.
- 3 Финн В.К. Своевременные замечания о ДСМ-методе автоматического порождения гипотез – 32 с.
- 4 ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания. Сост. О.М.Аншаков, Е.Ф.Фабрикантова. - М.: ЛИБРОКОМ, 2009. - 432 с.
- 5 Poletaykin A., Kostenko A. Use of jsm-method to find regularities of the road accidents rising // Автомобильный транспорт: сб. науч. трудов. Харьков : ХНАДУ, 2015. Вып. 36. С. 13–21
- 6 Правила дорожного движения Российской Федерации 2019.
- 7 Милль Дж. Ст., Система логики силлогистической и индуктивной. Изложение принципов доказательства в связи с методами научного исследования, 1914 г., Изд. Г. А. Лемана, с. 355.
- 8 Автоматическое порождение гипотез в интеллектуальных системах / Сост. Е.С. Панкратова, В.К. Финн; Под. общ. ред. В.К. Финна. — М.: ЛИБРОКОМ, 2009. – 528 с.
- 9 Михеенкова М.А. Принципы и логические средства интеллектуального анализа социологических данных: автореферат диссертации на соискание ученой степени доктора технических наук: 05.13.17.- М.:, 2011. — 51 с.: ил.
- 10 ДСМ-метод // [https://ru.wikipedia.org/wiki/ДСМ-метод#Описание\\_метода](https://ru.wikipedia.org/wiki/ДСМ-метод#Описание_метода) [13 декабря 2018]

11 Что такое Python // <https://pythonz.net/promo/> [25 августа 2019]

12 Узнайте все о Firebird // [https://www.firebirdnews.org/docs/fb2min\\_ru.html](https://www.firebirdnews.org/docs/fb2min_ru.html) [25 августа 2019]

13 Показатели состояния безопасности дорожного движения // <http://stat.gibdd.ru/> [25 августа 2019]