Оглавление

[1. Введение 8](#_Toc327289708)

[1.1. Актуальность темы и проблематика исследования 8](#_Toc327289709)

[1.2. Исходные данные 9](#_Toc327289710)

[1.3. Постановка задачи 9](#_Toc327289711)

[1.4. Обзор литературы 10](#_Toc327289712)

[1.4.1. Методы на основе анализа движущихся объектов 11](#_Toc327289713)

[1.4.2. Комбинированные методы 13](#_Toc327289714)

[1.4.3. Методы на основе аггрегации данных 14](#_Toc327289715)

[2. Исследовательская часть 20](#_Toc327289716)

[2.1. Анализ исходных данных 20](#_Toc327289717)

[2.2. Выбор верхнего порогового значения для скоростей 25](#_Toc327289718)

[2.3. Квантование данных по времени 25](#_Toc327289719)

[2.4. Обработка дублирующихся отсчетов 26](#_Toc327289720)

[2.5. Сглаживание временных радов после квантования 28](#_Toc327289721)

[2.5.1. Метод скользящего среднего. 28](#_Toc327289722)

[2.5.2. Метод наименьших квадратов 29](#_Toc327289723)

[2.5.3.Медианный фильтр 29](#_Toc327289724)

[3. Конструкторская часть 33](#_Toc327289725)

[3.1. Описание математической модели 33](#_Toc327289726)

[3.2. Классификация дней недели 33](#_Toc327289727)

[3.3. Расчет коэффициента влияния 35](#_Toc327289728)

[3.4. Расстояние между днями 36](#_Toc327289729)

[3.5. Восстановление пропущенных отсчетов 38](#_Toc327289730)

[3.6. Реализация математической модели 41](#_Toc327289731)

[3.7. Параметры математической модели 42](#_Toc327289732)

[3.7.1. Параметр количества похожих ситуаций – r 42](#_Toc327289733)

[3.7.2. Пороговое значение коэффициента влияния 42](#_Toc327289734)

[3.6.3. Пороговое значение отклонения скорости от среднего значения 43](#_Toc327289735)

[3.7.4. Нижнее пороговое значение для скорости 43](#_Toc327289736)

[3.7.5. Замена среднего значения 43](#_Toc327289737)

[3.8. Анализ полученных результатов 44](#_Toc327289738)

[4. Технологическая часть 48](#_Toc327289739)

[4.1. Определение набора ключевых критериев качества разрабатываемой системы 48](#_Toc327289740)

[4.2. Функциональное тестирование модулей 49](#_Toc327289741)

[4.2.1. Тестирование алгоритма расчета коэффициента влияния 49](#_Toc327289742)

[4.2.2. Тестирование алгоритма восстановления данных. 53](#_Toc327289743)

[4.3. Эффективность и модифицируемость системы 56](#_Toc327289744)

[4.4. Требования к переносимости системы 59](#_Toc327289745)

[4.5. Выводы 60](#_Toc327289746)

[5. Организационно-экономическая часть 62](#_Toc327289747)

[5.1 Задание к организационно-экономической части дипломного проекта 62](#_Toc327289748)

[5.2 Планирование работ по теме. Анализ процесса и трудоемкости 62](#_Toc327289749)

[5.3 Сетевая модель проекта 64](#_Toc327289750)

[5.4 Сетевой график проекта 68](#_Toc327289751)

[5.5 Календарный график выполнения проекта 69](#_Toc327289752)

[5.6 Анализ структуры затрат проекта 71](#_Toc327289753)

[5.6.1. Затраты на выплату исполнителям заработной платы 71](#_Toc327289754)

[5.6.2. Расходы на дополнительную заработную плату 72](#_Toc327289755)

[5.6.3. Отчисления с заработной платы 73](#_Toc327289756)

[5.6.4. Затраты на расходные материалы 73](#_Toc327289757)

[5.6.6. Затраты, связанные с амортизацией оборудования 75](#_Toc327289758)

[5.6.7. Накладные расходы 76](#_Toc327289759)

[5.6.8. Смета затрат на реализацию алгоритма 76](#_Toc327289760)

[5.7. Затраты на внедрение программного продукта (результата дипломного проекта). 77](#_Toc327289761)

[5.8. Планирование цены и прогнозирование прибыли. 78](#_Toc327289762)

[5.9. Расчет окупаемости проекта. 80](#_Toc327289763)

[5.10. Выводы 81](#_Toc327289764)

[6. Промышленная экология и безопасность 83](#_Toc327289765)

[6.1. Анализ вредных и опасных факторов при работе с ПЭВМ 83](#_Toc327289766)

[6.1.2. Требования к микроклимату, содержанию аэроинов и вредных химических веществ в воздухе на рабочих местах, оборудованных ПЭВМ 86](#_Toc327289767)

[6.2. Расчет освещенности на рабочем месте 93](#_Toc327289768)

[6.3. Расчет снижения количества выхлопных газов автомобиля 96](#_Toc327289769)

[Заключение 100](#_Toc327289770)

[Литература 101](#_Toc327289771)

# 1. Введение

## 1.1. Актуальность темы и проблематика исследования

Проблема прогнозирования автомобильных пробок сейчас довольно часто поднимается в средствах массовой информации и является одной из ключевых для любого обладателя личного автомобиля.

Ряд организаций уже успешно реализовали алгоритмы по прогнозированию пробок. В России количество их мало, и нельзя однозначно выделить лидера в этой сфере. Ряд реализованных моделей показали себя не очень хорошо, в остальных случаях компании ограничились предсказанием средним или отображением статистики за некоторых период в прошлом.

Зарубежные компании разработали ряд математических моделей, которые позволяют предсказывать дорожные затруднения и хорошо себя показали при внедрении в устройства автомобильной навигации. Все зарубежные системы прогнозирования тесно связаны с системами сбора данных. Как правило, данные о состоянии транспортного потока снимаются с датчиков, которые размещены на дорогах, а также поступают с устройств автомобильной навигации.

В России не существует систем, которые обеспечивают регулярные измерения скорости на дорогах. Поэтому на данный момент невозможно перенести опыт зарубежных исследователей для предсказания скорости дорог в российских городах. Системы прогнозирования, которые опираются на данные, собранные лишь с устройств автомобильной навигации, должны решать как проблему предсказания скорости, так и проблему неполноты данных.

Задача прогнозирования скорости в условиях неполноты данных, таким образом, является открытой и представляет большой интерес для исследования.

## 1.2. Исходные данные

Исходными данными являются данные о структуре дорожного графа Санкт-Петербурга, а также данные о скорости на ребрах графа в различные моменты времени, полученные с устройств автомобильной навигации.

Ребра графа – это участки дороги, на которых нет пересечений с другими дорогами. Вершины графа – это перекрестки дорог. Смежными ребрами называются ребра графа, у которых есть общая вершина. То есть смежное ребро – это ребро, на которое можно попасть из текущего или с которого можно попасть на текущее ребро.

Дорога с двусторонним движением представлена как два различных ребра графа. Данные по ним независимы и собираются отдельно.

Для каждого ребра графа дан его уникальный номер и длина. Мы не имеем информации о физическом местоположении ребра графа.

Данные по скоростям на ребрах графа представлены в следующем формате. Каждая запись содержит 2 временных отсчета – время въезда на ребро и время выезда с этого ребра на одно их смежных. То есть если разницу во времени поделить на длину ребра, мы сможем получить данные о скорости на этом ребре с некоторой погрешностью. Погрешность обусловлена неточностью измерения времени въезда и выезда с ребра, а также погрешностью при определении местоположения автомобиля.

Данные представлены за полтора месяца – с 01.09.2010 по 15.10.2010.

## 1.3. Постановка задачи

Необходимо разработать и реализовать математическую модель для краткосрочного прогнозирования скоростей на дорогах. Под краткосрочным прогнозированием понимается предсказание скорости на ребре графа на 1 час вперед при известных скоростях на данный момент. Схожая задача может стоять перед разработчиками автомобильных навигаторов, которые должны проложить оптимальный по времени маршрут с учетом информации о пробках в данный момент времени.

Основными задачи работы являются:

1. Исследование и корректировка исходных данных;
2. Решение проблемы неполноты данных;
3. Разработка математической модели прогнозирования скорости;
4. Анализ полученных результатов и оценка эффективности системы.

На данный момент существующие системы прогнозирования пробок, внедренные российскими компаниями, опираются на метод экстраполяции среднего. Поэтому критерием эффективности системы прогнозирования будет являться увеличение точности прогноза скорости относительно прогнозирования средним значением.

## 1.4. Обзор литературы

На данный момент существует множество работ, посвященных анализу транспортного потока. Существует несколько принципиальных подходов к анализу дорожной ситуации. Классификация методов приведена на рисунке 1.1.

Все методы можно разделить на три группы:

1. Методы на основе анализа индивидуальных объектов потока – в качестве исходных данных используется информация о траектории движения отдельных автомобилей. Внешние факторы, такие как время, погодные условия и прочие, не учитываются.
2. Методы на основе аггрегированных данных – в качестве исходных данных используются данные о скоростях движения автомобилей на конкретных участках дорожного графа в различные отсчеты времени. В данном классе методов не уделяется внимание анализу поведения конкретных объектов потока, анализируется лишь его общее состояние.
3. Методы, использующие сочетание индивидуального анализа и аггрегации данных.

Рис. 1.1. Классификация методов анализа транспортных потоков

### 1.4.1. Методы на основе анализа движущихся объектов

Основой метода служит анализ траектории движущихся объектов. Данных класс методов не является адаптивным. Предполагается, что траектории объектов известны заранее и не изменятся в течение поездки.

Примером алгоритмов данной категории являются алгоритмы обнаружения движущихся кластеров [13, 16].

Общий принцип основан на нахождении кластеров, состоящих из наборов краткосрочных траекторий различных автомобилей. Кластеры являются по сути наборами данных объектов, которые потенциально двигаются по одной и той же траектории – то есть потенциально являются источниками дорожных затруднений.

Существую модели, основанные на кластеризации траектории объектов целиком [13], и модели, в которых траектории движения объектов разбиваются на отрезки и производится кластеризация отрезков [16]. В результате обнаруживаются схожие «субтраектории» автомобилей. Второй метод имеет преимущество перед первым, так как позволяет обнаруживать высокую концентрацию автомобилей, даже когда их траектории совпадают лишь частично. Как правило, траектории автомобилей намного длиннее, чем отрезки пути, на которых возможно появление пробки. Поэтому разумнее искать сходства на коротких участках пути, нежели искать совпадения траекторий в целом.

Другие методы обнаружения автомобильных пробок основаны на выявлении областей плотного движения объектов. Один из таких методов позволяет обнаружить перемещающиеся кластеры движущихся объектов из базы данных[18]. Отличие данного подхода от вышеописанных состоит в том, что кластеры остаются одни и те же, но может меняться их положение или состав.

Обычно такие кластеры имеют место в пробках, которые образованы в результате появления препятствия на дороге, замедляющего общую скорость потока. Некоторые участники движения попадают в пробку сразу после появления препятствия, и быстро покидают ее, некоторые же попадают в пробку спустя некоторое время после появления затруднения, и остаются там на некоторое время. Таким образом, состав кластера, который описывает пробку, изменяется, в то время как сам кластер остается одним и тем же.

Аналогичный алгоритм решает задачу обнаружения областей движущихся объектов, в которых сохраняется высокий уровень плотности транспортного потока в течение длительного времени [15]. Алгоритм также позволяет осуществлять краткосрочный прогноз состояния транспортного потока на основе текущего состояния.

### 1.4.2. Комбинированные методы

Комбинированные методы анализа дорожных ситуаций сочетают в себе черты методов на основе индивидуального анализа движущихся объектов и методов на основе аггрегированния данных. Примером комбинированного метода является алгоритм, позволяющий обнаружить *хотспоты* - дорожные участки, на которых часто возникают дорожные затруднения, и которые оказывают сильное влияние на состояние смежных участков дорог [17].

Обнаружение хотспотов является важной задачей, так как подобные участки дороги встречаются практически в каждом большом городе. Они регулярно блокируют дорожное движение в часы-пик, и участники дорожного движения вынуждены много времени тратить на ожидание в пробках. Обнаружение хотспотов также осуществляется через анализы траекторий движения отдельных автомобилей.

Еще один метод, относящийся к категории комбинированных методов, посвящен изучению поведения объектов транспортного потока [19]. С помощью регрессионного анализа на основе исторических данных о движении объектов строится математическая модель, описывающая поведение автомобиля в транспортном потоке. Она позволяет обнаружить «шаблонные» поведения автомобилей в транспортном потоке. Далее при прогнозировании поведения конкретного объекта вычисляется, к какому типу принадлежит поведения данного автомобиля, и строится прогноз траектории и скорости его движения. Данные метод является адаптивным и позволяет строить краткосрочный прогноз скорости на участках дорог. Данный метод эффективен в ситуациях, когда поведения автомобиля соответствует математической модели, на которую опирается алгоритм. Модель не подстраивается под изменение факторов, не описанных в модели, и не строит достоверного прогноза для автомобилей с нетипичным поведением.

### 1.4.3. Методы на основе аггрегации данных

Существует большое количество методов данного класса. Их общая черта – тесное взаимодействие с системами сбора данных. Один из методов, к примеру, получал данные от 900 датчиков, размещенных на различных участках дорог [20].

В рамках данной работы используются данные, формат которых не позволяет строить модель на основе индивидуального анализа объектов движения. Именно поэтому модели на основе аггрегации данных представляют особый интерес в качестве прототипов для построения математической модели прогнозирования скорости.

Примером успешной реализации и внедрения системы по предсказанию автомобильных затруднений является система, разработанная фирмой Microsoft [14].

Система использует алгоритм, основанный на анализе ключевых участков дорожного графа - хотспотов. Термин *хотспот* в рамках этого метода имеет тот же смысл, что и в методе, описанном ранее.

Метод требует более подробного рассмотрения, так как не просто позволил статистически предсказывать дорожные затруднения, но и получил хорошую оценку при реальном использовании в устройствах автомобильной навигации.

Обычно для проверки эффективности алгоритма разрабатывается некоторая метрика, которая служит показателем качества. Алгоритм может показать высокую точность предсказания при использовании определенной метрики, но это далеко не означает, что он применим на практике. Метрика может не учитывать множество параметров, которые оказываются решающими в реальности. Примером может служить случай, когда алгоритм дает точные предсказания для выходного дня и плохо предсказывает будни. Или предсказания точны для коротких ребер дорожного графа, а для длинных ребер оказываются не столь эффективными. Также возможны ситуации, когда алгоритм, хорошо зарекомендовавший себя на графе дорог одного города, дает неточный прогноз для другого дорожного графа.

Именно поэтому метод предсказания дорожной ситуации невозможно гарантированно отнести к разряду эффективных, пока не будет разработана качественная метрика для оценки предсказаний, либо пока алгоритм не повысит эффективность прокладки автомобильных маршрутов с учетом пробок.

Именно поэтому рассматриваемая система прогнозирования, прошедшая испытания в реальных условиях, может считаться успешной и требует особого внимания.

Первым этапом данного алгоритма является обнаружение областей-хотспотов. Для сбора данных использовались мобильные устройства автомобилистов, которые отсылали на сервер данные о скорости движения в различные моменты времени. Также в качестве хотспотов были приняты общеизвестные «узкие места» дорожного графа Сиэтла, где проводились испытания.



Рис. 1.2. Визуализация ключевых проблемных мест на дорожном графе Сиэтла.

В основу алгоритма легла гипотеза о том, что прогнозируя состояния хотспотов, можно эффективно предсказывать состояние всего графа в целом.

Для прогнозирования загруженности ключевых мест проводился анализ множества факторов, таких как погодные условия, начало школьных каникул, праздники и спортивные мероприятия. Также использовались данные о дорожных происшествиях, которые собирались с помощью обратной связи от пользователей навигационных устройств.

В зависимости от состояния каждого из факторов выбиралось наиболее вероятное состояние хотспотного участка.

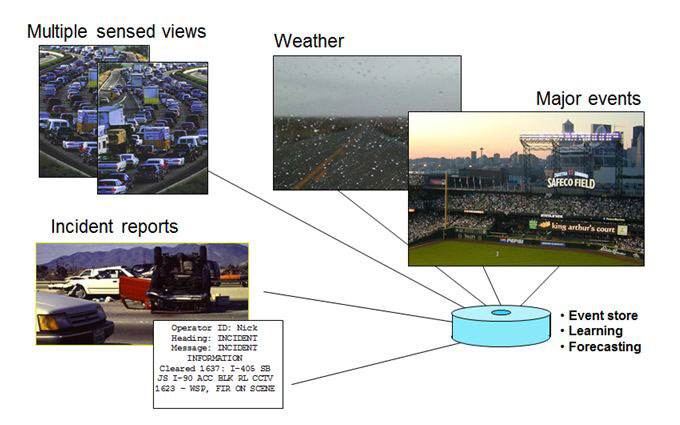


Рис. 1.3. Учет различных факторов при построении прогноза.

После того, как удалось получить прогноз состояния узких мест графа, необходимо было получить прогноз для остальных участков. Для этого каждое ребро, не являющееся хотспотом, было соотнесено с некоторым ключевым участком. На основе состояния ключевого участка делался вывод о состоянии в соответствующих ему ребрах графа.

Помимо сопоставления участкам дорожного графа ключевых мест было проведено исследование, которое находило взаимосвязи между ключевыми местами. Часто затруднение движения на одном из ключевых участков вызывало затруднения на другом.



Рис. 1.4. Ключевое место 15 (отмечено звездочкой) оказывало влияние на ключевые участки с номерами 17, 11 и 4.

Таким образом, была получена система, которая на основе состояния набора ключевых факторов прогнозировала состояние дорог. Причем для каждого ребра учитывалось состояние абсолютно всех факторов и прогнозирование проводилось по одному и тому же алгоритму для каждого ребра.

Для каждого ребра, помимо основного алгоритма прогнозирования, был рассчитан индивидуальный метод. Были выявлены факторы, которые сильно влияют на конкретный хотспотный участок, остальные факторы для этого участка просто не учитывались. Выяснилось, что для большинства ключевых мест достаточно учитывать только несколько факторов, причем для каждого хотспота количество основных факторов было небольшим (3-5) по сравнению с общим количеством факторов (несколько десятков).

Индивидуальный прогноз использовался для указания, насколько правдиво предсказание, полученное с помощью общего алгоритма.

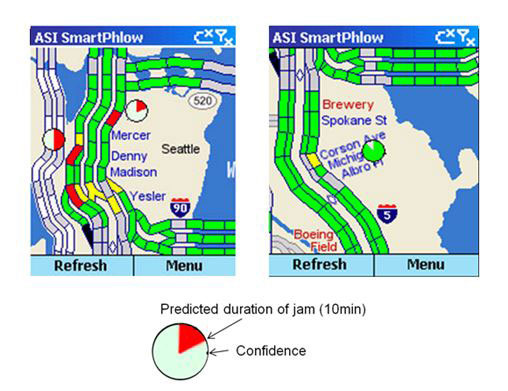


Рис. 1.5. Использование условного знака для обозначения достоверности прогноза.

Таблица 1.1.

Точность предсказания состояния ключевых участков дорожного графа.

|  |  |  |  |
| --- | --- | --- | --- |
| Номер ключевого места | Точность  (Свободно, загружено) | Номер ключевого места | Точность  (Свободно, загружено) |
| 0 | 0.83, 0.92 | 11 | 0.76, 0.93 |
| 1 | 0.75, 0.93 | 12 | 0.65, 0.98 |
| 2 | 0.78, 0.91 | 13 | 0.70, 0.95 |
| 3 | 0.83, 0.90 | 14 | 0.83, 0.95 |
| 4 | 0.87, 0.96 | 15 | 0.80, 0.84 |
| 5 | 0.73, 0.94 | 16 | 0.73, 0.86 |
| 6 | 0.65, 0.95 | 17 | 0.78, 0.92 |
| 7 | 0.84, 0.97 | 18 | 0.76, 0.86 |
| 8 | 0.85, 0.92 | 19 | 0.68. 0.94 |
| 9 | 0.81, 0.96 | 20 | 0.82, 0.96 |
| 10 | 0.71, 0.91 | 21 | 0.86, 0.94 |

Существуют также ряд исследований по прогнозированию дорожной ситуации, которые были проведены в рамках конкурса Яндекс.Математика 2010. В статьях победителей были изложены основные подходы к прогнозированию пробок и обзор результатов применения различных методик прогнозирования [2], [7]. Несмотря на то, что алгоритмы, представленные в работах, не были опробованы при прогнозировании пробок в реальных условиях, статьи содержат несколько интересных выводов и их можно рассматривать как первое приближение в разработке алгоритма предсказания дорожной ситуации.

# 2. Исследовательская часть

## 2.1. Анализ исходных данных

В качестве исходных данных мы имеем набор отсчетов – значения скорости в различные моменты времени для ребер графа.

Временные ряды, образованные этими отсчетами, неэквидистантны. Также во многих исследованиях, посвященных прогнозированию дорожной ситуации, отмечается наличие ощутимого тренда в изменении скоростей – со временем общая скорость на дорогах падает. Это значит, что дорожная ситуация с каждым годом ухудшается и средняя скорость движения в часы пик изменяется в меньшую сторону. Следовательно, не имеет смысла говорить о стационарности имеющихся временных рядов – предметная область такова, что они по сути нестационарны. Также мы не можем говорить о независимости значений временного ряда от реализации - она принципиально одна – следовательно, процесс также будем считать неэргодичным.

Первым этапом анализа является разделение ребер на группы по их длине. Большая часть ребер (около 70%) имеет длину меньше 250м. Следовательно, время движения по этим ребрам может составлять несколько секунд. Поскольку нам неизвестна погрешность измерения данных, можно было предположить, что временные ряды для коротких ребер содержат слишком сильную составляющую случайной погрешности и их невозможно использовать для дальнейших исследований.

Рис. 2.1. Распределение ребер графа по длине

В каждой группе ребер были посчитаны средние значения скоростей для разного времени суток.

Рис. 2.2. Средняя скорость в течение суток для ребер длиной 0 – 250м.

Рис. 2.3. Средняя скорость для ребер длиной 250 - 500м за сутки.

Рис. 2.4. Средняя скорость для ребер длиной 500м – 1км в течение суток.

Рис. 2.5. Средняя скорость для ребер длиной 1 – 2км в течение суток.

Рис. 2.6. Средняя скорость для ребер длиной 2 – 4км в течение суток.

Рис. 2.7. Средние значения скоростей для ребер длиной более 4км в течение суток.

Сделаем выводы о некоторых особенностях входных данных на основе полученных графиков.

* Погрешность содержат как короткие, так и длинные ребра, о чем свидетельствуют выбросы амплитуды на рисунках 2.6 и 2.7. Тем не менее, для коротких ребер наблюдаются минимумы скорости в утренние и вечерние часы пик – в результате усреднения была получена типичная картина изменения скорости.
* Скорости для коротких ребер в среднем меньше, чем скорости для длинных ребер. Видимо, длинные ребра соответствуют большим трассам – аналог МКАДа в Москве.
* Временной ряд содержит скорости более 1000км/ч. Следовательно, существует значительная погрешность измерения скорости. К сожалению, невозможно полностью исключить отклонения скоростей в меньшую сторону – низкие значения скоростей могут быть связаны как с ошибкой измерения, так и с затором на дороге в данный момент времени. Отклонения в большую сторону можно частично отсеять.

## 2.2. Выбор верхнего порогового значения для скоростей

Назначим верхнюю границу показания скорости в 150км/ч. Все записи со значениями скорости больше 150км/ч будут отбрасываться.

Дополнительно скорректируем записи с высокой скоростью. Если на дороге отсутствует затор, то выбор скорости движения зависит только от предпочтений водителя. Скорость на свободной дороге теоретически может иметь абсолютно любое значение. На практике комфортная скорость движения составляет от 60км/ч и выше. На больших трассах максимальная разрешенная скорость может быть выше 60км/ч. В такой ситуации водители предпочитают двигаться со средней скоростью потока, которая будет близка к максимально допустимой.

Исходя из данных соображений, проведем корректировку высоких скоростей по формуле

Vкорр. = min{Vисх., max{ 60км/ч, Vразр. }}, где

Vисх. – исходная скорость в записи,

Vразр. – максимальная разрешенная скорость на ребре

Vкорр. – полученная скорректированная скорость.

Верхний предел скорости совпадает с максимально разрешенной скоростью на ребре. Исключения составляют ребра с низкой разрешенной скоростью. Чаще всего водители нарушают правила и превышают допустимую скорость. Если на ребре с допустимой скоростью 40км/ч все автомобили движутся со скоростью 40км/ч и меньше – это может говорить о пробке на данном участке. Если же скорость составляет 60км/ч – то движение свободно. Поскольку эти показания характеризуют принципиально различные дорожные ситуации, мы будем их различать (вместо приведения обоих записей к максимально допустимой скорости в 40км/ч).

## 2.3. Квантование данных по времени

Разница во времени между появлениями соседних отсчетов временного ряда обусловлена методикой измерения. Данные о скорости на дороге поступают с устройств, установленных на автомобилях. С одной стороны это приводит к значительным потерям отсчетов временного ряда – если в данный момент на дороге нет ни одного автомобиля со специальным устройством, данных по этому ребру нет. С другой стороны, мы можем получить сразу несколько значений скорости для одного и того же значения времени (либо разница во времени настолько мала, что ей можно пренебречь).

Нужно определить, что считается отсчетом. Существующие системы, занимающиеся анализом и прогнозированием скоростей на дорогах, проводят дискретизацию данных по времени с некоторым шагом. К примеру, сервис Яндекс.Пробки предоставляют данные о пробках с интервалом в 4 минуты, а с ростом эффективности системы измерения, сократили интервал до 2х минут.

Чем лучше устроена система сбора информации, тем меньший шаг дискретизации можно выбрать. Верхний предел шага дискретизации определяется здравым смыслом. Например, при шаге в 30 минут нет смысла пользоваться прогнозом дорожной ситуации, если время в пути составит меньше часа.

С учетом вышеизложенного был выбран интервал в 15 минут. В дальнейшем, при наличии большего количества исходных данных, ожидается сокращение шага дискретизации по времени.

Все отсчеты округляются вниз до ближайшего 15-минутного значения. Например, отсчет в 18-39 будет рассматриваться как отсчет в 18-30.

## 2.4. Обработка дублирующихся отсчетов

При дискретизации данных по времени несколько записей о скорости приходятся на один и тот же отсчет. Рассматривались два варианта решения:

1. Результирующая скорость – среднее значение скоростей для данного интервала.
2. Результирующая скорость – медианное значение скоростей.

Был принят второй вариант обработки данных. Он позволяет при достаточном количестве исходных данных отбрасывать нехарактерные поведения автомобилей. На дороге могут одновременно ехать водители со скоростью 60км/ч (средний водитель в потоке), 20км/ч (неисправный автомобиль) и 120км/ч (водитель - «лихач»). Понятно, что оценку дорожной ситуации стоит оценивать по значению в 60км/ч – в данном случае средней скорости потока.

Рис. 2.8. Скорость в течение суток для ребра №15219 за 14.09.10 – вторник.

Рис. 2.9. График изменения скоростей для ребра №12625 за 14.09.2010 – вторник.

После квантования мы получили графики изменения скоростей на различных ребрах в течение суток. Наблюдаются высокие скорости движения ранним утром и ночью. В утренние и вечерние часы (около 9:00 и 19:00) скорость движения падает. На сильно загруженных ребрах днем (12:00 – 19:00) может не наблюдаться повышение скорости. На дорогах с относительно невысокой транспортной загруженностью днем скорость немного выше, чем в утренние и вечерние часы пик.

На графиках заметны выбросы значений скорости – резкое повышение скорости днем или пилообразное изменение графика скорости ночью. Пики такого рода на графиках вероятнее всего обусловлены поведением отдельных автомобилей в потоке и не характеризуют общую картину загруженности ребра.

## 2.5. Сглаживание временных радов после квантования

Для устранения выбросов скорости, которые не характеризуют ситуацию на ребре в целом, проведем процедуру сглаживания временных рядов.

Рассмотрим несколько распространенных алгоритмов и выберем наиболее подходящий для нашей задачи.

### **2.5.1. Метод скользящего среднего.**

Заключается в замене фактических значений членов ряда средним арифметическим значений нескольких ближайших к нему членов. Набор усредняемых значений образует так называемое окно скольжения. Член, значение которого заменяется на среднее по окну, занимает в окне срединное положение.

Метод хорошо подходит для рядов с частыми, но незначительными по амплитуде выбросами значений. На графиках показано, что исходные данные содержат большие выбросы скорости. Как только выброс будет попадать в окно скольжения, среднее значение будет сильно искажаться и сглаженная функция будет сильно отличаться от исходных данных.

### **2.5.2. Метод наименьших квадратов**

Используется в том случае, когда требуется функцию, заданную в *m* точках (*xj*, *yj*), *j* = 1, 2,…, *m*, приблизить многочленом степени *n* < *m*:

|  |
| --- |
| http://www.graphicon.ru/oldgr/grafor/gr_help/chapter_5.files/equat050401.gif. |

Нахождение многочлена, который опишет зависимость скорости на ребре от времени, само по себе является довольно сложной задачей. Поэтому этот метод также не является наилучшим решением.

### **2.5.3.Медианный фильтр**

Медианный фильтр является наиболее устойчивым к влиянию помех большой амплитуды. Поэтому мы будет применять именно его для решения поставленной задачи.

Используем модифицированный алгоритм медианного фильтра (аналогичный алгоритм применяется в функции medsmooth программы Mathcad).

Обозначим исходный временной ряд через vy. Размер окна фильтра обозначим через n. Алгоритм сглаживания функции состоит из 4х этапов:

1. Формируется сглаженный вектор vy’. Его*i*-ый элемент дается выражением: *vy'*i = median(vyi-(n-1)/2,...vyi,...vyi+(n-1)/2).
2. Затем вычисляются остатки: **vr = vy - vy'**.
3. Вектор остатков, **vr**, сглаживается с использованием процедуры, описанной на шаге 1. Это дает сглаженный вектор остатков **vr'**.
4. Возвращается сумма из этих двух сглаженных векторов:  **vy' + vr'**.

Необходимо выбрать величину окна фильтра.

Рис. 2.9. Сравнение результатов обработки временного ряда медианным сглаживанием с разными размерами окна фильтрации.

Результирующая функция при n=3 исключает выбросы с большой амплитудой, но сохраняет незначительные колебания значений скорости.

Сглаживание при n=5 и n=7 дает схожий результат – из ряда исключаются большие выбросы значений скорости, а незначительные колебания сглаживаются. Поскольку принципиальных различий в результате нет, выберем параметр, при котором функция сглаживания будет работать быстрее. Таким образом, для дальнейших исследований будем использовать сглаживающую функцию с параметром n=5.

Нужно убедиться, что процедура сглаживания не испортила основные характеристики временного ряда. В рамках поставленной задачи график среднесуточной скорости должен удовлетворять следующим требованиям:

* Минимумы скорости должны находиться в районе 9:00 и 19:00.
* Ночью скорость заметно возрастает
* В период с 9:00 до 19:00 наблюдается небольшой подъем значений

На рис. 2.10. представлен график изменения средней скорости по всем ребрам в течение суток до сглаживания и фильтрации данных. График отличается высокой частотой колебания амплитуды скорости и содержит выбросы. По исследованиям аналитического центра Яндекс.Пробки средняя загруженность дорог изменяется в течение суток плавно и имеет явные экстремумы во время утреннего и вечернего часа пик (рис. 2.11). Процедура сглаживания по сути является процедурой фильтрации шумов, которые присутствуют в исходных данных.

Рис. 2.10. Средние значения скоростей для всех ребер до квантования и предварительной обработки.

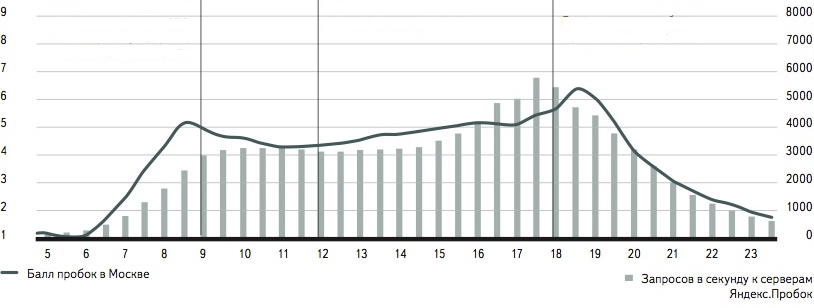


Рис. 2.11. Изменение уровня загруженности дорог в течение суток по данным аналитического центра Яндекс.Пробки.

Рис. 2.12. Среднесуточная скорость для всех ребер после процедуры сглаживания.

На графике маркерами отмечены локальные минимумы. Форма полученной функции сохраняет признаки, характерные для среднесуточной скорости движения автомобилей.

Следовательно, мы не внесли во временной ряд изменения, которые значительно ухудшили качество данных и осуществили фильтрацию шумов.

# 3. Конструкторская часть

## 3.1. Описание математической модели

Для прогнозирования дорожной ситуации был выбран метод ближайших соседей. Суть его состоит в следующем: при предсказании скорости в какой-то момент времени мы должны найти ситуацию в прошлом, максимально похожую на данную, и продублировать показания скорости с ситуации в прошлом на настоящий момент. Этот метод позволит предсказывать пробки, даже когда мы не имеем полной информации о факторах, влияющих на дорожную ситуацию. Мы оцениваем некоторые тенденции и копируем их без подробного анализа причин, вызвавших затруднения. Поскольку мы не можем предсказать аварии, ремонт дороги или хотя бы собрать информацию о такого рода данных, данный метод кажется особенно привлекательным и уже хорошо себя зарекомендовал в ряде исследований.

Этапы прогнозирования:

1. Определяем, к какой категории дней недели относится рассматриваемый день
2. Для всех дней в прошлом из той же категории рассчитываем величину расстояния до рассматриваемого дня в момент t
3. Выбираем r дней с наименьшим расстоянием
4. Для всех отсчетов t + 15j вычисляем среднее значение скорости в выбранные дни
5. Полученные скорости и будут прогнозом

## 3.2. Классификация дней недели

При прогнозировании скорости необходимо учитывать множество факторов. Одним из них является день недели, для которого ведется прогнозирование.

В будние дни изменение скорости транспортного потока резко отличается от изменения скорости в выходной день.

Все дни недели разделим на 5 категорий:

1. Будни – это дни с понедельника по четверг. Исследования показывают, что в будни пики и спады скоростей примерно совпадают по времени. Поэтому среды, как правило, похожи на четверги и вторники. При прогнозировании скорости в среду мы будем рассматривать как данные по средам, так и по остальным дням из этой категории.
2. День перед выходными (как правило пятница) – дорожная ситуация по пятницам как правило похожа на остальные дни. Но она отличается от остальных, например, более ранним окончанием рабочего у большого количества людей – то есть несовпадением вечернего часа пик с остальными буднями. Также летом в пятницу люди начинают активно выезжать за город, что вызывает большие пробки практически в течение целого дня. Важно учесть, что при выходном дне посреди рабочей недели, предыдущий день будет очень похож на обычную пятницу, так как сопровождается сокращенным рабочим днем.
3. Первый день выходных (как правило, суббота) – отличается низкой загруженностью дорог, а в летнее время наличие утренних пробок из-за поездок большого количества людей за город.
4. Выходные (чаще всего отсутствуют) – в случае праздничных дней, длящихся более 2х дней, второй, третий и т.д. дни выходных дороги несильно загружены и пробок практически нет.
5. Последний день выходных (обычно воскресение) – то же самое, что и первый день выходных, но с учетом вечерних пробок.

В обычной неделе есть будни, пятница (день перед выходными), суббота (первый день выходных), воскресение (последний день выходных).

В случае, когда на неделю выпадает какой-либо государственный праздник, сокращенный день переносится на четверг и далее следует три дня выходных, стоит заново классифицировать дни недели. Понедельник, вторник и среда будут относиться к будням. Четверг - день перед выходными. Пятница – первый день выходных, далее суббота (выходной) и воскресение – последний день выходных.

Рис. 3.1. Деление дней недели на категории.

Если в середине недели выходной (например, вторник нерабочий) – то его будем относить к категории выходные. В случае отсутствия данных можем прогнозировать его на основе данных из категории 3 и 5 (как субботу и воскресение). Понедельник же в этой ситуации будет считаться днем перед выходными.

## 3.3. Расчет коэффициента влияния

Для прогнозирования скорости на ребре дорожного графа стоит учитывать изменение скоростей на смежных ребрах. Часто дорожное затруднение на смежном ребре является причиной дорожного затруднения на исследуемом ребре и наоборот.

Необходимо определить критерий, показывающий, насколько сильно изменение скорости на смежном ребре связано с изменением скорости на исследуемом ребре.

Таким критерием будет является коэффициент влияния смежного ребра на основное.

Значение коэффициента влияния будет равно

ρ = M / N, где

N – количество отсчетов, в которых на основном ребре наблюдается отклонение скорости от среднего значения на величину ∆V,

M – количество отсчетов, в которых отклонение скорости от среднего на величину ∆V отличается сразу и на основном, и на смежном ребре.

В процессе проведения исследований величина ∆V выбиралась как 30% от значения скорости в конкретный момент времени.

Таким образом, коэффициент влияния показывает, насколько часто отклонения скорости от нормы на основном ребре сопровождается отклонением скорости на смежном ребре.

Из определения следует, что коэффициент влияния ребра на самого себя равен 1.

## 3.4. Расстояние между днями

Выбранная математическая модель требует введения понятия расстояния между днями.

Искать ближайшую дорожную ситуацию мы будем среди дней той же категории, что и день недели, для которого ведется прогнозирование. Также при прогнозировании для времени t мы будем смотреть ситуации только в эти же моменты t в другие дни.

Для расчета расстояния будут учитываться скорости как в момент времени t, для которого ведется прогнозирование, так и скорости в моменты времени t-15, t-30 и т.д. Скорости в выбранные моменты времени будут сравниваться как для исследуемого ребра, так и для смежных с ним ребер.

Разница скоростей в отсчетах будет браться с учетом коэффициента влияния ребра на исследуемое. Если коэффициент влияния смежного ребра близок к 0, это означает, что изменение скорости на этом ребре не оказывает влияния на скорость основного ребра. И разница скоростей этого смежного ребра не должна значительно увеличивать расстояние между ситуациями. И наоборот, если смежное ребро оказывает сильное влияние на основное, то разница скоростей на таком ребре должна сильно увеличивать расстояние.

Формула для расчета расстояния между днями:

Dd – расстояние между ситуацией, которая считается текущей на момент прогнозирования, и ситуацией в некоторый день d;

ρi - коэффициент влияния ребра i на искомое;

Vi,td – скорость на ребре i в момент времени t дня d;

M – количество рассматриваемых ребер;

t – момент времени, который считается текущим;

N – количество отсчетов, которые нужно учитывать при прогнозировании.

Пример расчета:

Вычислим расстояние между двумя пятницами для ребра 1. Ребра 2 и 3 являются смежными к нему. Расчет проводится для 18:00. N равно двум, то есть будем учитывать 2 отсчета в прошлом. Для 18:00 это будут моменты 17:45 и 17:30.

Таблица 3.1. Пример данных

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| № ребра | ρ | Данные для первого дня | | | Данные для второго дня | | |
| 17:30 | 17:45 | 18:00 | 17:30 | 17:45 | 18:00 |
| 1 | 1 | 45 км/ч | 50 км/ч | 75 км/ч | 48 км/ч | 54 км/ч | 79 км/ч |
| 2 | 0.9 | 25 км/ч | 35 км/ч | 45 км/ч | 20 км/ч | 30 км/ч | 40 км/ч |
| 3 | 0.1 | 25 км/ч | 35 км/ч | 45 км/ч | 15 км/ч | 25 км/ч | 35 км/ч |

Хотя скорости на ребре №3 отличались друг от друга сильнее, чем скорости других ребер, они внесли меньший вклад в общее расстояние между ситуациями из-за маленького значения коэффициента влияния ρ.

## 3.5. Восстановление пропущенных отсчетов

Исходные данные содержат большое количество пропущенных отсчетов. В среднем известен лишь каждый 10й отсчет, то есть одно значение скорости на 2.5 часа.

Рис. 3.2. Пример ребра с высокой плотностью известных отсчетов.

Рис. 3.3. Пример ребра с низкой плотностью известных отсчетов.

Формула, введенная ранее для вычисления расстояния между днями, не учитывает ситуации, когда какой-либо из отсчетов отсутствует.

Было принято решение восстановить значения скоростей в отсчетах, содержащих пробелы.

Для восстановления был использован алгоритм интерполяции среднего.

Если у нас известны два отсчета, а между ними присутствуют 1 или 2 неизвестных, то их восстанавливаем, считая, что скорость на этом интервале изменялась линейно.

Если между двумя известными отсчетами больше двух пропусков, то все временные интервалы, отличающиеся от известных более чем на 30 минут, мы заполняем ближайшими средними значениями. Оставшиеся по краям находим, как описано выше.

Таблица 3.2. Пример восстановления данных.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Исходные данные (км/ч) | Известные средние | Восстановленные данные | Комментарий |
| 17:15 | 64 | 57 | 64 | Известно |
| 17:30 | - | - | 64.3 | Линейно изменяется |
| 17:45 | - | 65 | 64.6 | Линейно изменяется |
| 18:00 | - | - | 65 | Ближайшее известное среднее |
| 18:15 | - | 72 | 72 | Известно среднее |
| 18:30 | - | 71 | 77 | Изменяется линейно |
| 18:45 | - | - | 82 | Изменяется линейно |
| 19:00 | 87 | 82 | 87 | Известно |

Следует отметить, что данный алгоритм восстановления хорошо работает на временных рядах с низкой частотой колебания амплитуд, к которым относятся временные ряды, описывающие скорость движения автомобильного потока. Для временных рядов с высокой частотой колебания амплитуд погрешность восстановления велика. Данный метод стоит использовать как базовый при испытаниях математической модели. В дальнейшем рекомендуется разработать алгоритм восстановления с более низким значением ошибки.

Рис. 3.3. Пример восстановленных данных

На рис. 3.3. показан график временного ряда на ребре после восстановления пропущенных отсчетов. К характерным особенностям временного ряда после восстановления можно отнести:

1. Длительные интервалы ряда с одинаковой скоростью – результат замены пропущенных отсчетов средним.
2. Значительная разность скоростей на стыке участков, где данные по скорости известны, и участков с пропущенными отсчетами.

## 3.6. Реализация математической модели

Выбранная математическая модель была реализована на языке программирования python 3. Он был выбран, так как обладает простым синтаксисом и имеет возможность подключения библиотек языка R, специализированного на обработке статистических данных. Минус выбранного решения – низкое быстродействие модели. Но поскольку на данном этапе работы важно было получить адекватную математическую модель, быстродействием можно было пренебречь.

Прогнозирование скоростей для какого-либо дня состоит из следующих этапов:

1. Получение информации о ребре и смежных с ним ребер.
2. Расчет средних значений для полученных ребер.
3. Расчет коэффициентов «влияния» для всех ребер.
4. Восстановление данных.
5. Прогнозирование для выбранного временного участка.
6. Вычисление ошибки прогнозирования.

Для каждого пункта был разработан самостоятельный модуль. Все промежуточные значения записываются в текстовые файлы. То есть мы можем впоследствии модифицировать каждую итерацию независимо от других. При корректировке какого-то этапа можно использовать данные полученные на предыдущих этапах без повторных расчетов. Также такая схема позволяет писать каждый модуль на своем языке и создавать распределенные системы на вычислительных кластерах. Подробнее особенности реализации будут рассмотрены в разделе 4.

## 3.7. Параметры математической модели

### 3.7.1. Параметр количества похожих ситуаций – r

Параметром r будем далее называть число схожих ситуаций, которые следует учитывать при прогнозировании скорости. Мы считаем «расстояния» для каждой ситуации в прошлом для искомой, выбираем r наиболее похожих ситуаций. Изначально этот параметр был выбран как 5. В работе, схожей с данной[1], отмечается, что значения коэффициентов следует выбирать от 1 до 5. Для улиц с большим количеством данных эффективнее прогнозирование с меньшими значениями коэффициента r и наоборот.

### 3.7.2. Пороговое значение коэффициента влияния

Для каждого из смежных ребер мы рассчитываем коэффициент влияния его на основную дорогу. Если смежное ребро не влияет на основное, то использование данных по этому ребру может ухудшить предсказание, внося дополнительные помехи. Следовательно ребра с очень низким коэффициентом влияния не следует учитывать совсем. Была введена величина influence\_boundary, которая означает пороговое значение для коэффициента влияния. Все ребра, не превышающие эту границу, не будут учитываться при окончательных расчетах. Изначально значение было установлено как 0.1.

### 3.7.3. Пороговое значение отклонения скорости от среднего значения

При расчете коэффициента влияния мы ищем ситуации, когда скорость на дороге отличается от среднего значения. Разницу вычисляем в процентах. Введем еще один параметр, который будет указывать минимальный процент, на который должны отличаться две скорости, чтобы отклонение считалось значимым. Изначально значение было установлено в 30%.

### 3.7.4. Нижнее пороговое значение для скорости

Входные данные, как отмечалось ранее, отличаются большими погрешностями. И часто можно встретить значение скоростей в 0км/ч или в 350 км/ч. Так что при расчете коэффициентов было решено отсечь хотя нижние значения скоростей и введен параметр порога скорости. Если скорость на участке ниже пороговой, то он не принимался в рассмотрение при подсчете коэффициентов влияния. В этой реализации модели равен 1 км/ч.

### 3.7.5. Замена среднего значения

При восстановлении данных требовалось некоторые неизвестные значения заменить ближайшим средним. Для будних дней почти всегда есть хотя бы несколько известных отсчетов, поэтому для них всегда возможно восстановление средним. Для пятниц и выходных данные часто совсем отсутствуют. Тогда в качестве среднего выбиралось среднее значение из другой категории дней. Было принята следующая система переходов.

1. Если нет данных за пятницу – берем данные для будних дней.
2. Если нет данных за субботу – применяем среднее значение скорости в воскресение. Если нет и этих данных, выбираем в качестве опорных данные в будние дни.
3. Если нет данных за воскресение – заменяем средними значениями скорости в субботу. Если нет и этих данных, выбираем в качестве опорных данные в будние дни.

В общем случае для восстановления данных достаточно иметь несколько отсчетов за будние дни. Но понятно, что восстановленные данные будут содержать большую погрешность.

## 3.8. Анализ полученных результатов

Прогнозирование проводилось для последней недели известных данных, то есть для чисел 11.10.2010 по 17.10.2010, то есть с понедельника по воскресение. Историческими данными являлись записи для 01.09.2010 по 10.10.2010.

Для прогнозирования скорости использовались исторические данные, а также данные о текущем состоянии дорожной ситуации.

Рассмотрим характерные особенности спрогнозированных данных на примере.

На рис. 3.4. представлен пример спрогнозированных на 30 минут вперед данных. Прогнозирование велось для 15.10.2010, то есть пятницы.

Рис. 3.4. Сравнение реальных и спрогнозированных данных. Прогноз на 30 минут вперед.

Участок с 6:00 до 18:00 содержит одинаковые прогнозируемые значения по скорости. Это означает, что для данного ребра мало отсчетов для пятницы – следовательно, данные восстанавливались средним значением. При отсутствии исторических данных прогноз будет соответствовать среднему значению скорости в этот момент времени.

Для отрезка, приходящегося на вечерний час пик (18:00 – 21:00), прогноз верно сработал в сторону снижения скорости – модель предсказала образование дорожного затруднения на этом отрезке времени.

Причем стоит отметить, что прогнозируемое начало затруднения опережает реальное начало пробки. Если бы прогнозируемая пробка отставала от реальной, предсказание теряло бы смысл – если пробка действительно началась, то водитель и так самостоятельно поймет, что ситуация на дороге будет только ухудшаться в ближайшие полчаса.

Рис. 3.5. Сравнение ошибок прогнозирования для ребер с различной плотностью исходных данных

Величина средней ошибки сильно варьируется и зависит в основном от количества исходных данных для прогнозируемого ребра.

На рисунке 3.5. показаны два типичных случая:

1. Ошибка прогнозирования невелика и превышает по точности прогнозирование средним значением.
2. Ошибка прогнозирования велика и уступает по точности прогнозирование средним значением.

Случай 1 наиболее распространенный и является типичным для ребер с большим количеством входных данных.

Случай 2 наблюдается на ребрах с очень маленьким набором известных отсчетов. При отсутствии исходных данных для ребра лучше себя показывает экстраполяция по среднему значению.

Была сделана попытка рассчитать погрешности для разных значений параметра r (сколько похожих дорог учитывается при прогнозировании). Но на небольшом количестве данных тенденция совершенно неявная и приводит то к увеличению ошибки прогнозирования, то к уменьшению.

Рис. 3.6. Сравнение ошибок прогнозирования при различных r.

В качестве оптимального значения было принято r=3.

Полученная математическая модель решила поставленную задачу. Она превосходит по точности прогнозирование средним. В среднем плотность известных отсчетов составляла один отсчет на 2.5 часа, то есть 9 из 10 отсчетов обычно отсутствуют. Для ребер, на которых плотность исторических данных была ниже, модель показала себя хуже, чем прогнозирование средним.

Абсолютная средняя погрешность для дорог составляла около 10 км/ч.

# 4. Технологическая часть

## 4.1. Определение набора ключевых критериев качества разрабатываемой системы

Рассмотрим основные критерии качества, предъявляемые к программному обеспечению:

* Функциональность – наличие всех функций, удовлетворяющих требованиям пользователя
* Надежность – способность поддерживать заданный уровень работоспособности
* Практичность – наличие средств, дающих пользователю возможность изучения, понимания и использования системы.
* Эффективность (производительность и масштабируемость) – способность системы функционировать согласно требованиям, предъявляемым к производительности и масштабируемости.
* Модифицируемость – способность системы к модификации для улучшения или адаптации ее работы
* Переносимость – система может быть адаптирована для работы в различных средах

В рамках данной работы интерес представляют требования к функциональности, масштабируемости, модифицируемости и переносимости.

К системе не предъявляются требования по надежности, так как она является прототипом для исследований и не предполагает внедрения в системы реального времени. Также предъявляются минимальные требования к практичности - все программные модули сопровождаются технической документацией.

## 4.2. Функциональное тестирование модулей

### 4.2.1. Тестирование алгоритма расчета коэффициента влияния

Одним из этапов выбранной математической модели прогнозирования пробок является расчет коэффициента влияния смежного ребра дорожного графа на ребро, для которого ведется прогнозирование.

Расчет коэффициента проводится следующим образом.

Проведем квантование временных рядов с интервалом 15 минут и посчитаем среднее значение скорости в эти интервалы для рассматриваемых ребер дорожного графа. Затем пройдем по всем известным отсчетам и посчитаем количество ситуаций, когда

1. На основном ребре наблюдается отклонение скорости от среднего значения на какую-то определенную величину.
2. Отклонение скорости от среднего отличается сразу и на основном и на смежном ребре.

Вторая величина, поделенная на первую, и будет являться коэффициентом влияния. Другими словами, она покажет, насколько часто отклонения скорости от нормы на основном ребре сопровождается отклонением скорости на смежном ребре.

Из определения следует, что коэффициент влияния ребра самого на себя равен 1.

Если на основном ребре не наблюдается отклонений от среднего значения, то значение коэффициента влияния определить невозможно. Такие случаи имеют место, когда данные содержат большое количество пропущенных отсчетов, а также на ребрах, где не наблюдались дорожные затруднения.

Проведем ряд испытаний для проверки алгоритма расчета. Для этого смоделируем изменения скорости на основном и смежном ребрах с помощью различных функций. Для них рассчитаем ожидаемое значение коэффициента и сравним с результатами работы алгоритма.

Тест 1.

Vосн(t) = 60;

Vсмежн(t) = 60;

Vсреднее = 60;

Полученное значение коэффициента влияния = -1.

Полученное значение верно, так как скорости на основном ребре ни разу не отклонились от среднего значения – невозможно определить величину коэффициента влияния.

Тест 2.

Vосн(t) = 10;

Vсмежн(t) = 60 + sin(hours \* 4 + minutes / 15) \* 60;

Vсреднее = 60;

Вычислим предполагаемое значение коэффициента влияния.

Скорость на основном ребре всегда отклоняется от среднего значения больше, чем на 30%.

Следовательно, коэффициент влияния равен относительному количеству отсчетов, в которых скорость на смежном ребре отличается от среднего больше, чем на 30%.

Поскольку график является синусоидой, которая колеблется около среднего значения, при пороговом значении в 30% искомое количество отсчетов будет равно:

1 – arcsin(0.3) / (π/2) = 0.806

Сгенерируем данные по приведенным формулам и рассчитаем значение коэффициента влияния.

Рис. 4.1. Временные ряды, изучаемые в тесте №2.

Полученное значение коэффициента = 0.78.

Полученное значение немного отличается от теоретического, так как осуществлялось квантования данных, и функция изменения скорости в течение суток содержит нецелое количество периодов.

Тест 3.

Vосн(t) = 60 - sin(hours \* 4+minutes/15) \* 60;

Vсмежн(t) = 60 + sin(hours \* 4+minutes/15) \* 60;

Vсреднее = 60;

Зададим функции изменения скорости на ребрах как синусоиды, которые колеблются в противофазе.

Отклонения от среднего значения на обеих синусоидах должно происходить в одних и тех же отсчетах. При этом рубеж в 30% с большей вероятностью пройдет отсчет, скорость в котором на данный момент ниже. Таким образом, ожидаемое значение коэффициента влияния должно принять значение 0.8÷1.

Рис. 4.2. Графики скоростей для теста №3.

Полученное значение коэффициента влияния 0.956 подтверждает теоретические выводы.

Таблица 4.1.

Результаты тестирования алгоритма расчета коэффициента влияния

|  |  |  |  |
| --- | --- | --- | --- |
| № теста | Входные параметры | Расчетное значение коэффициента влияния | Полученное значение коэффициента влияния |
| 1 | Vосн(t) = 60;  Vсмежн(t) = 60;  Vсреднее = 60; | -1 | -1 |
| 2 | Vосн(t) = 10;  Vсмежн(t) = 60 + sin(hours \* 4 + minutes / 15) \* 60;  Vсреднее = 60; | 0.806 | 0.78 |
| 3 | Vосн(t) = 60 - sin(hours \* 4+minutes/15) \* 60;  Vсмежн(t) = 60 + sin(hours \* 4+minutes/15) \* 60;  Vсреднее = 60; | 0.8 ÷ 1 | 0.956 |

### 4.2.2. Тестирование алгоритма восстановления данных.

Временные ряды содержат большое количество пробелов. Поскольку выбранная математическая модель прогнозирования предполагает дискретный временной ряд без пробелов, необходимо осуществлять восстановление отсчетов.

Восстановление осуществляется методом интерполяции среднего. Если не хватает 3 и менее отсчетов – линейно интерполируем пропущенные данные по крайним известным отсчетам. Если пропущен длинный ряд, делаем опорные точки по среднему значению и также интерполируем по известным отсчетам с учетом опорных значений.

Для тестирования алгоритма восстановления данных сгенерируем данные с вероятностью потери отсчета 20%.

Тест 1.

Функция временного ряда: V = 60 + sin(hours) \* 20

Данная функция не характерна для изменения скоростей автомобиля. Поэтому ожидается, что восстановленные данные будут содержать значительную величину средней ошибки.

Рис. 4.3. Восстановление функции V=60 + 20sin(hours).

Средняя величина ошибки составила 25.58км/ч. Для вероятности пропуска отсчета в 20% это очень большая величина погрешности. Как и предполагалось, данный алгоритм восстановления не предназначен для функций с частыми колебаниями амплитуды.

Тест №2

Возьмем функцию из теста 1 и увеличим период колебаний. Ожидается значительное уменьшение значения средней ошибки.

V = 60 + 20sin(hours / 10)

Рис. 4.4. Восстановление функции V = 60 + 20sin(hours/10)

Среднее значение погрешности восстановления данных составила 7.34 км/ч.

Таблица 4.2.

Результаты тестирования алгоритма восстановления данных

|  |  |  |  |
| --- | --- | --- | --- |
| № теста | Входные параметры | Ожидаемая величина ошибки | Полученная величина ошибки |
| 1 | V = 60 + sin(hours) \* 20  Pпот = 0.2 | > 10 км/ч | 25.58 км/ч |
| 2 | V = 60 + 20sin(hours / 10)  Pпот = 0.2 | <10 км/ч | 7.34 км/ч |

Выбранный алгоритм восстановления данных хорошо работает на временных рядах, которые не содержат резких колебаний скорости. Этому условию соответствует временной ряд, описывающий изменения скорости автомобильного потока на дороге.

Рис. 4.5. Пример восстановления данных для некоторого ребра.

При анализе графиков восстановленных данных невозможно определить погрешность результатов в силу того, что реальные данные отсутствуют. Можно лишь проанализировать общие тенденции изменения скорости на ребре и сделать предположения об успешности алгоритма восстановления.

На графике мы наблюдаем резкое снижение скорости на ребре в утренние часы и увеличение скорости движения в вечерние часы. Это характерное поведение функции скорости на ребре в рабочий день – в часы пик образуются и рассасываются автомобильные заторы. Восстановленные данные подтверждают теоретические предположения о форме функции скорости, следовательно данный алгоритм возможно применять при решении поставленной задачи.

## 4.3. Эффективность и модифицируемость системы

Рис. 4.6. Структура системы прогнозирования

Система прогнозирования транспортной загруженности дорог разбита на 6 модулей, изображенных на рисунке 6.

Модули вызываются последовательно из программы или консоли и работают независимо друг от друга. Каждый модуль осуществляет чтение данных из файла. Выходные данные также записываются в текстовый файл. Для каждого файла установлены требования по его формату. Выходной файл одного модуля может служить в качестве источника входных данных для другого модуля.

Рассмотрим основные сценарии работы с системой и решения, обеспечивающие эффективности работы с системой в предложенных сценариях.

Во всех сценариях источником стимула является программист-исследователь, а артефактом сама система анализа и прогнозирования данных.

Таблица 4.3.

Анализ критериев качества системы

|  |  |  |
| --- | --- | --- |
| **Стимул** | **Реакция** | **Количественная мера реакции** |
| Тестирование результатов работы модулей | Получение промежуточных результатов расчетов | Результаты работы каждого модуля записываются в текстовые файлы. Таким образом, все промежуточные результаты сохраняются после завершения прогнозирования. Исследователь не тратит время на отладку.  Поскольку модули работают независимо друг от друга, можно выполнять запуск только одного модуля и подавать на вход тестовые наборы данных через запись в соответствующий входной файл. |
| Прогнозирование с использованием математической модели восстановления данных, отличной от исходной | Необходимо иметь возможность замены любого из промежуточных модулей системы | Система разбита на модули. Каждый модуль осуществляет чтение данных из файла и, в свою очередь, записывает результаты работы в текстовый файл.  При реализации нового промежуточного модуля необходимо лишь соблюсти интерфейс входных и выходных данных. А также заменить вызов старого модуля на новый в программе, запускающей моделирование. |
| Необходимо сравнить результаты работы системы при использовании двух различных модулей восстановления данных | Система должна предоставлять возможность моделирования при двух наборах составных модулей и дальнейшего сравнения результатов. | Все модули принимают на вход в качестве параметра название директории, в которую записываются промежуточные и конечные результаты расчетов.  Структура размещения промежуточных результатов внутри выбранной директории фиксирована.  Это позволяет проводить моделирование несколько раз при различных условиях и затем получить результаты всех этапов. |
| Необходимо реализовать модуль предварительной обработки данных на языке C++. | Система должна предоставлять возможность использования модулей, написанных на различных языках программирования. | Объектно-ориентированная природа Python, являясь мощным средством  структурирования программного кода многократного пользования, кроме  того, делает этот язык идеальным инструментом поддержки сценариев для объектно-ориентированных языков, таких как C++ и Java. Например, при наличии соответствующего связующего программного кода, программы на  языке Python могут использовать механизм наследования от классов,  реализованных на C++, Java и С#. |

## 4.4. Требования к переносимости системы

Стандартная реализация языка Python написана на переносимом ANSI С, благодаря чему он компилируется и работает практически на всех основных платформах. Например, программы на языке Python могут выполняться на самом широком спектре устройств, начиная от наладонных компьютеров (PDA) и заканчивая суперкомпьютерами. Список основных операционных систем и устройств, где можно использовать Python:

• Операционные системы Linux и UNIX

• Microsoft Windows и DOS (все современные версии)

• Mac OS (обе разновидности: OS X и Classic)

• BeOS, OS/2, VMS и QNX

• Системы реального времени, такие как VxWorks

• Суперкомпьютеры Cray и ЭВМ производства компании IBM

• Наладонные компьютеры, работающие под управлением PalmOS, PocketPC

или Linux

• Сотовые телефоны, работающие под управлением операционных систем

Symbian и Windows Mobile

• Игровые консоли и iPod

Помимо самого интерпретатора языка в составе Python распространяется стандартная библиотека модулей, которая также реализована переносимым способом. Кроме того, программы на языке Python компилируются в переносимый байт-код, который одинаково хорошо работает на любых платформах, где установлена совместимая версия Python.

Все это означает, что программы на языке Python, использующие основные возможности языка и стандартные библиотеки, будут работать одинаково и в Linux, и в Windows, и в любых других операционных системах, где установлен интерпретатор Python. В большинстве реализаций Python под определенные операционные системы имеется также поддержка специфических механизмов этих систем (например, поддержка СОМ в Windows), но ядро языка Python и библиотеки работают совершенно одинаково в любой системе. Как уже говорилось выше, Python включает в себя средства создания графического интерфейса Tk GUI под названием tkinter, что позволяет программам на языке Python создавать графический интерфейс, совместимый со всеми основными графическими платформами без индивидуальной программной настройки.

## 4.5. Выводы

Основными критериями качества системы являются требования к функциональности. Результаты работы системы на наборах тестовых данных совпали с расчетными значениями.

Также система удовлетворяет предъявляемым требованиям по функциональности, масштабируемости, модифицируемости и переносимости. Требования к надежности системы не предъявляются, так как она не предназначена для эксплуатации в системах реального времени.

# 5. Организационно-экономическая часть

## 5.1 Задание к организационно-экономической части дипломного проекта

Задачи организационно – экономической части дипломного проекта:

* 1. Определение структуры (этапов) работ по созданию ПО;
  2. Расчет трудоемкости проекта
  3. Определение численности исполнителей;
  4. Построение сетевого графика выполнения проекта;
  5. Разработка календарного графика работ;
  6. Определение структуры затрат на разработку проекта;
  7. Анализ сегмента рынка, разрабатываемого ПО;
  8. Оценка экономической целесообразности реализации проекта.

## 5.2 Планирование работ по теме. Анализ процесса и трудоемкости

Разработка системы по прогнозированию дорожной ситуации ведется двумя работниками:

1. Ведущий разработчик.
2. Разработчик.

Общие затраты труда на разработку и внедрение проекта определяются по формуле (1):

,

где - затраты труда на выполнение i*-* гоэтапа проекта.

В таблице 5.1 приведен перечень этапов проекта. Значения трудоемкости определены на основании аналогичных работ, проводимых ранее.

Таблица 5.1.

Перечень работ проекта

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Трудоемкость**  **(человеко-дни)** | |
| **Ведущий разработчик** | **Разработчик** |
| **Этапы работ** | 1. Разработка ТЗ | 3 | 3 |
| 2. Разработка архитектуры системы статистического анализа. | 7 | 7 |
| 3. Разработка модуля обработки первичных данных. | 0 | 2 |
| 4. Разработка модуля корректировки скоростей. | 0 | 2 |
| 5. Разработка модуля сглаживания. | 0 | 2 |
| 6. Тестирование и анализ модулей предварительной обработки данных. | 3 | 3 |
| 7. Разработка блока управления данными. | 0 | 5 |
| 8. Разработка модуля восстановления данных. | 3 | 0 |
| 9. Разработка модуля прогнозирования данных. | 3 | 0 |
| 10. Разработка модуля оценки ошибки прогнозирования. | 0 | 3 |
| 11. Отладка и тестирование системы статистического анализа. | 5 | 5 |
| 12. Выбор моделей для восстановления данных. | 5 | 5 |
| 13. Изучение моделей восстановления данных на обучающей выборке. | 7 | 7 |
| 14. Выбор моделей для прогнозирования. | 5 | 5 |
| 15. Изучение моделей прогнозирования на обучающей выборке. | 7 | 7 |
|  | 16. Анализ полученных результатов. | 2 | 2 |
|  | 17. Составление итогового отчета по результатам исследований. | 3 | 3 |

Общие затраты труда на разработку = 114 человеко-дней.

## 5.3 Сетевая модель проекта

Сетевой график устанавливает взаимосвязь между всеми работами проекта и позволяет окончательно определить длительность как отдельных этапов, так и всего проекта в целом.

Построение сетевого графика предполагает использование метода сетевого планирования, на базе которого разрабатывается информационно-динамическая модель процесса выполнения проекта. Построение сетевой модели включает оценку степени детализации комплекса работ, определения логической связи между отдельными работами и временные характеристики выполнения этапов проекта

В сетевой модели выделяют события и работы. В качестве событий, например, принимают факты начала проекта, окончания разработки отдельных модулей, интерфейсов, выполнения отладки и т.п. Все события нумеруются по порядку от исходного к завершающему.

Графическое отображение сетевой модели (сетевой график) содержит окружности, отображающие основные события проекта, и векторы, соединяющие эти окружности и определяющие необходимость выполнения соответствующих работ. Реальные работы изображаются сплошной линией, фиктивные - штриховой, а работы, лежащие на критическом пути - линией двойной толщины.

Окружности, как показано на рис. 5.1, разделены на четыре сектора, в которых показаны:

* номер данного события (в нижнем секторе);
* значение раннего срока наступления текущего события (в левом секторе);
* значение резерва времени текущего события (в верхнем секторе);
* значение позднего срока наступления события (в правом секторе).



Рис. 5.1 обозначение основных элементов сетевого графика.

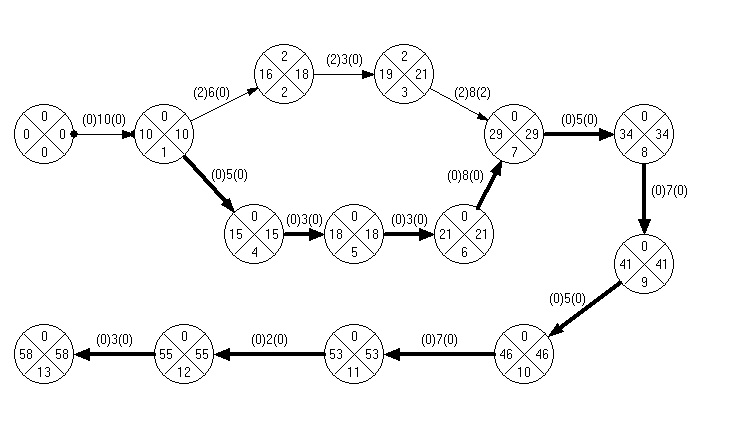
На данном рисунке: - номер события, - ранний срок наступления события i, - поздний срок наступления события i, - резерв времени события i, - продолжительность работы i-j, - полный резерв времени работы i-j, - свободный резерв времени работы i-j.

Таблица 5.2

Основные события и работы проекта

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **N** | **Событие** | **Код работы** | **Работа** | **t,** | |
| **чел/часы** | **чел/дни** |
| 0 | Начало работ | 0-1 | Разработка ТЗ.  Разработка архитектуры системы анализа. | 160 | 20 |
| 1 | Разработка архитектуры системы статистического анализа. | 1-2 | Разработка модулей первичной обработки данных. | 48 | 6 |
|  |  | 1-4 | Разработка модуля управления данными | 40 | 5 |
| 2 | Разработка модуля первичной обработки данных. | 2-3 | Разработка модуля оценки ошибки. | 24 | 3 |
| 3 | Разработка модуля оценки ошибки. | 3-7 | Тестирование и отладка модулей системы. | 64 | 8 |
| 4 | Разработка модуля управления данными | 4-5 | Разработка модуля восстановления данных | 24 | 3 |
| 5 | Разработка модуля восстановления данных. | 5-6 | Разработка модуля прогнозирования. | 24 | 3 |
| 6 | Разработка модуля прогнозирования. | 6-7 | Отладка и тестирование системы статистического анализа. | 64 | 8 |
| 7 | Получение отлаженной системы анализа данных | 7-8 | Выбор моделей восстановления данных. | 80 | 10 |
| 8 | Получение моделей восстановления данных | 8-9 | Исследование моделей восстановления данных на обучающей выборке | 112 | 14 |
| 9 | Выбор наилучшей модели восстановления данных. | 9-10 | Выбор моделей прогнозирования данных | 80 | 10 |
| 10 | Получение моделей прогнозирования | 10-11 | Исследование моделей прогнозирования | 112 | 14 |
| 11 | Выбор наилучшей модели прогнозирования | 11-12 | Анализ полученных результатов | 32 | 4 |
| 12 | Получение результатов исследований | 12-13 | Написание технической документации | 48 | 6 |
| 13 | Окончание работ |  |  |  |  |

## 5.4 Сетевой график проекта

****

Критический путь - максимальный путь от исходного события (0) до завершения проекта.

Lкр = L0-1 + L1-4 + L4-5 + L5-6 + L6-7 + L7-8 + L8-9 + L9-10 + L10-11 + L11-12 + L12-13 = **58** дней.

Рис. 5.2. Сетевой график выполнения дипломного проекта.

## 5.5 Календарный график выполнения проекта

Для иллюстрации последовательности проводимых работ проекта применяют ленточный график (календарно-сетевой график, диаграмму Гантта). На рисунке 5.3 показан график Гантта.



Продолжительность работ, календарные дни

Рис.5.3. Диаграмма Гантта.

## 5.6 Анализ структуры затрат проекта

Затраты на выполнение проекта состоят из прямых затрат (на заработную плату исполнителям, затрат на закупку или аренду оборудования, затрат на организацию рабочих мест), и косвенных затрат (на т.н. накладные расходы) (2).

 (2)

где *СЗАРП -* заработная плата исполнителей*, СОБ -* затраты на обеспечение необходимым оборудованием, *СОРГ -* затраты на организацию рабочих мест, *СНАКЛ* - накладные расходы.

### 5.6.1. Затраты на выплату исполнителям заработной платы

Затраты линейно связаны с трудоемкостью и определяется следующим соотношением:

 (3)

где *СЗ.ОСН -* основная заработная плата*, СЗ.ДОП -* дополнительная заработная плата*, СЗ.ОТЧ -* отчисление с заработной платы*.*

Расчет основной заработной платы (оплаты труда непосредственных исполнителей) следует проводить по «дневной» оплате труда на основе данных по окладам и графику занятости исполнителей (4):

 (4)

где *ТЗАН -* число дней, отработанных исполнителем проекта, *ОДН -* дневной оклад исполнителя. При 8-и часовом рабочем дне он рассчитывается по соотношению 5:

 (5)

где *ОМЕС* - месячный оклад, *FM  -* месячный фонд рабочего времени.

С учетом налога на доходы физических лиц размер месячного оклада увеличивается, что отражено в формуле 6:

 (6)

где *О* - «чистый» оклад, который получен из информации кадровых агентств, *НДФЛ -* налог на доходы с физических лиц (13%).

Перечень исполнителей, их месячные и дневные оклады, а также время участия в проекте и рассчитанная основная заработная плата для каждого исполнителя показаны в таблице 5.3.

Таблица 5.3.

Затраты на основную заработную плату

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Исполнители** | **Трудоемкость,**  **человеко-дни** | **Оклад, руб./мес.** | **Дн. оклад, руб.** | **Затраты по з/п, руб.** |
| 1 | Ведущий разработчик | 58 | 60 000 | 2850 | 165300 |
| 2 | Разработчик | 56 | 40 000 | 1900 | 106400 |
| **ИТОГО** | | | | | **271700** |

### 5.6.2. Расходы на дополнительную заработную плату

Расходы учитывают все выплаты непосредственным исполнителям за время, не проработанное на производстве, но предусмотренное законодательством, в том числе: оплата очередных отпусков, компенсация за недоиспользованный отпуск, и др. Величина этих выплат составляет 20% от размера основной заработной платы:

 (7)

Для рассматриваемой разработки составляет **54 340** рублей.

### 5.6.3. Отчисления с заработной платы

Отчисления с заработной платы состоят в настоящее время в уплате страховых взносов в государственные внебюджетные фонды. С 01.01.2012 общий тариф страховых взносов составляет 30% (до 01.01.2012 - 34%).

Согласно налоговому кодексу РФ, применяются ставки налога для отчисления в пенсионный фонд РФ, фонд социального страхования, федеральный фонд обязательного медицинского страхования.

Отчисления с заработной платы составят:

 (8)

где *НСОЦ* – включает в себя следующие отчисления:

* **страховые взносы на обязательное пенсионное страхование** – **22%;**
* **страховые взносы на обязательное социальное страхование на случай временной нетрудоспособности и в связи с материнством - 2,9%;**
* **страховые взносы в ФФОМС - 5,1%.**

Исходя из этого, *НСОЦ* составит 30%. Следовательно, С*З.ОТЧ* = **97812** рубля. И тогда затраты на выплату заработной платы для рассматриваемой разработки равны С*зарп* = 271700 + 54340 + 97812 = 423 852 рубля.

### 5.6.4. Затраты на расходные материалы

Затраты определяются перечнем минимально необходимых для организации работ расходных материалов. Данный перечень приведен в таблице 5.4.

Таблица 5.4.

Затраты на расходные материалы

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№**  **п/п** | **Наименование** | **Цена, руб.** | **Кол-во, шт.** | **Сумма, руб.** |
| 1 | Бумага офисная CANON (500л, А4) | 280 | 1 | 280 |
| 2 | Оплата услуг интернет (1 Гб трафика) | 550 | 1 | 550 |
| Итого затраты на расходные материалы С*расх* | | | | **830** |

**5.6.5. Затраты на закупку или аренду оборудования**

Разрабатываемая система построена с применением механизма распределенных параллельных вычислений, что позволяет гибко развертывать ее, используя как существующее, так и новое оборудование. Вычислительная мощность оборудования определяется только тем, какая глубина анализа и производительность необходима. Минимальная рекомендуемая конфигурация аппаратных средств это персональный компьютер, отвечающий требованиям, приведенным в таблице 5.5.

Таблица 5.5.

Минимальная конфигурация аппаратных средств

| Параметр | Значение |
| --- | --- |
| Процессор | Intel Core I5 или более производительный |
| Оперативная память | 8 ГБ и более |
| Жесткий диск | 2 Тб и более |
| Среда исполнения | Eclipse netbeans |
| Операционная система | Ubuntu |
| Монитор | Любой модели |

Рыночная стоимость данного оборудования составляет 16000 рублей.

Для разработки необходимо купить два персональных компьютера по цене 16000 рублей за штуку (по одному для каждого разработчика) и один принтер по цене 10000 рублей (для изготовления сопроводительной документации). Следовательно, суммарные затраты на однократное приобретение оборудования, необходимого для разработки системы, составляют:

Таким образом, суммарные расходы на закупку оборудования составляют **ССО = 42 000 рублей.**

### 5.6.6. Затраты, связанные с амортизацией оборудования

Затраты определяются по формуле (9):

где – затраты на приобретение i-го средства производства, – время использования i-го средства производства в днях, – полный ожидаемый срок эксплуатации i-го средства производства в днях.

В ходе работ используется следующее оборудование (см. таблицу 5.6.):

Таблица 5.6.

Затраты на амортизацию оборудования

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Оборудование** | **Стоимость, руб.** | **Кол-во, шт.** | **Норм. срок службы,**  **дни** | **Эксплуат., дни** | **Затраты, руб.** |
| Аппаратные средства | 32 000 | 1 | 765 | 114 | 5 000 |
| Принтер Epson AcuLaser C900 | 6 500 | 1 | 520 | 114 | 1 425 |
| **ИТОГО:** | | | | | **6 425** |

Принимая полный срок эксплуатации приобретенных ПЭВМ 3 года (с учетом морального старения), а для прочего оборудования – от 2 до 6 лет, и учитывая, что в году приблизительно 255 рабочих дней, получаем:

**=6 425 рублей.**

### 5.6.7. Накладные расходы

Расходы, связанные с выполнением проекта, вычисляют, ориентируясь на расходы по основной заработной плате. Обычно они составляют от 60% до 100% расходов на основную заработную плату. Тогда:

 (10)

и **C*НАКЛ***  для рассматриваемого проекта будут составлять=0,6∙271 700=**163020 рублей.**

### 5.6.8. Смета затрат на реализацию алгоритма

Таблица 5.7.

Смета затрат на реализацию алгоритма

|  |  |  |
| --- | --- | --- |
| **№** | **Наименование статей** | **Затраты, руб.** |
| 1 | Основная заработная плата | 271700 |
| 2 | Дополнительная заработная плата | 54340 |
| 3 | Отчисления на страховые взносы | 97812 |
| 4 | Затраты на закупку оборудования | 42 000 |
| 5 | Затраты на расходные материалы | 830 |
| 6 | Амортизационные отчисления | 6425 |
| 7 | Накладные расходы | 161020 |
| **ИТОГО:** | | **634 127** |

Рис. 5.4. Круговая диаграмма затрат на выполнение дипломного проекта

## 5.7. Затраты на внедрение программного продукта (результата дипломного проекта).

Затраты на внедрение программного продукта (результата дипломного проекта) состоят из затрат на заработанную плату исполнителям со стороны фирмы-разработчика, затрат на закупку оборудования, необходимого для внедрения программного продукта, затрат на организацию рабочих мест и оборудование рабочего помещения и затрат на накладные расходы.

Затраты на внедрение определяются из соотношения:

,

где СВН.ЗАРП — заработанная плата исполнителям, участвующим во внедрении,

СВН.ОБ — затраты на обеспечение необходимым оборудованием,

СВН.ОРГ — затраты на организацию рабочих мест и помещений,

СВН.НАКЛ — накладные расходы.

Так как работы по внедрению могут проводиться на оборудовании, ранее установленном заказчиком, и на рабочих местах заказчика, то СВН.ОБ и СВН.ОРГ равны нулю. Расчет затрат на выплату заработной платы и накладные расходы следует проводить исходя из того, что время внедрения составляет 7 дней, и работами по внедрению будет заниматься один программист с окладом 40000 руб. Тогда

Учитывая все затраты в случае единичного внедрения ПО, общие затраты составят следующую сумму:

## 5.8. Планирование цены и прогнозирование прибыли.

Процесс разработки сложной программной продукции сопровождается, кроме решения чисто программных аспектов, необходимостью решения многих социальных и экономических проблем. Одна из экономических проблем — определение стоимости программного продукта.

На основе данных о затратах на разработку и внедрение, результатах прогнозирования объема продаж, следует определить стоимость одного комплекта программного обеспечения. Стоимость поставляемого на рынок ПО определяется частью стоимости разработки ПО, затрат на внедрение и прибыли фирмы — разработчика.

«Рыночную» стоимость ПО можно рассчитать, используя соотношение 11:

 (11)

где *ΔK -* часть стоимости разработки, приходящаяся на одну копию программы, *КВН* - стоимость внедрения программы, *DПРИБ* - процент прибыли, заложенный в стоимость.

Стоимость внедрения остается постоянной для каждой установки ПО, а частичная стоимость разработки, приходящаяся на каждый комплект ПО, определяется исходя из данных о планируемом объеме установок (12):

 (12)

где *К*- стоимость проекта, *NP* - планируемое число копий ПО, *HСТ* - ставка банковского процента по долгосрочным кредитам (более одного года).

В качестве процента по долгосрочным кредитам берем 18% годовых. Планируемое число копий составляет 30.

Тогда:

По результатам мониторинга рынка установим *К ПР* = 100 000 руб. и тогда с учетом данных о стоимости комплекта программы, стоимости установки и частичной стоимости разработки можно определить процент прибыли от одной реализации ПО:

 (13)

Сумму расчетной прибыли от продажи каждой установки ПО следует определять, используя следующее соотношение:

 (14)

где *ННДС* - процентная ставка налога на добавочную стоимость.

## 5.9. Расчет окупаемости проекта.

Рассчитаем, при каком количестве проданных экземпляров системы окупятся расходы на ее разработку.

При отсутствии кредитов на внедрение (ввиду незначительности величины затрат на внедрение и самого срока внедрения), взятии долгосрочного кредита со ставкой 18% годовых на разработку ПП и частичное погашения кредита, общий баланс составит суммы показанные в приведенной ниже таблице.

Таблица 5.8.

Общий баланс.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Период расчета, мес.** | **Число проданных копий, шт.** | **Баланс нач., руб.** | **Сумма продаж, руб.** | **Сумма погашения кредита, руб.** | **Факт. совокупн. прибыль** | **Факт. чистая прибыль** | **Баланс конечный, руб.** |
| 01.06.2012- 30.06.2012 | 2 | - | 200000 | 64069 | 135931 | 101079 | -550474 |
| 01.07.2012- 31.07.2012 | 4 | -514571 | 200000 | 34521 | 65479 | 55490,68 | -449395 |
| 01.08.2012- 31.08.2012 | 6 | -459081 | 200000 | 34521 | 65479 | 55490,68 | -348316 |
| 01.09.2012- 30.09.2012 | 8 | -403590 | 200000 | 34521 | 65479 | 55490,68 | -247237 |
| 01.10.2012- 31.10.2012 | 10 | -348099 | 100000 | 34521 | 65479 | 55490,68 | -146158 |
| 01.11.2012- 30.11.2012 | 12 | -292609 | 100000 | 34521 | 65479 | 55490,68 | -45079 |
| 01.12.2012- 31.12.2012 | 14 | -237118 | 100000 | 34521 | 65479 | 55490,68 | 56000 |

К началу 2013 года разработка программы окупится. Выплата кредита будет осуществляться ежемесячно в течение 1 года. Короткий срок окупаемости обусловлен тем, что программный продукт рассчитан на узкий круг покупателей, будет приобретен ими сразу или в течение полугода и сразу принесет ожидаемую прибыль.

## 5.10. Выводы

В данном разделе нами была определена трудоемкость проекта, число исполнителей проекта, необходимые затраты на разработку, внедрение, а также произведено планирование цены и прогнозирование прибыли. Значения рассчитанных экономических параметров для данного дипломного проекта представлены в таблице 5.8.

Таблица 5.8. Экономические параметры дипломного проекта

| **Обозначение** | **Название экономического параметра** | **Значение параметров** |
| --- | --- | --- |
|  | Трудоемкость проекта | 114 чел/дня |
|  | Количество исполнителей проекта | 2 чел |
|  | Затраты на заработную плату исполнителям | 423 852 руб. |
|  | Затраты на оборудование | 7275 руб. |
|  | Накладные расходы | 161 020 руб. |
|  | Затраты на внедрение | руб. |
|  | Общие затраты | 634127 руб. |
|  | Стоимость одной копии разработанного ПО | 100 000 руб. |
|  | Размер прибыли с одной копии | руб. |

Проект окупится после продажи 14 копий системы или спустя 7 месяцев после начала продаж. Прибыль после продажи каждой следующей копии составит 46600 руб.

Разрабатываемая система прогнозирования дорожной ситуации может быть использована как самостоятельный информационный сервис, так и внедрена в устройства автомобильной навигации по маршруту.

Также система может послужить прототипом для разработки системы статистического анализа в другой предметной области.

# 6. Промышленная экология и безопасность

## 6.1. Анализ вредных и опасных факторов при работе с ПЭВМ

Статистические исследования и расчеты в данной работе ведутся с активным использованием персональных электронно-вычислительных машин (ПЭВМ). Поэтому необходимо соблюдать санитарно-эпидемиологические требования к ПЭВМ и условиям труда, согласно СанПин 2.2.2/2.4.1340-03. Данные санитарные правила необходимы для предотвращения неблагоприятного влияния на здоровье человека вредных факторов производственной среды и трудового процесса с ПЭВМ.

В рамках данного дипломного проекта разработчику приходилось работать со следующим оборудованием:

* ПЭВМ;
* видеодисплейный терминал (ВДТ);
* клавиатура;
* манипулятор "мышь";
* принтер.

Оборудованию, указанному в списке соответствуют следующие опасные и вредные факторы:

* электромагнитное поле;
* статическое электричество;
* акустический шум;
* визуальные показатели ВДТ;
* мягкое рентгеновское излучение;
* пожарная опасность;
* электрический ток.

Воздействие вредного фактора на человека может привести к профессиональному заболеванию. Случай воздействия опасного фактора на работающего человека называется несчастным случаем.

**6.1.1Требования к ПЭВМ и ВДТ**

Временные допустимые уровни электромагнитных полей (ЭМП), создаваемых ПЭВМ, не должны превышать значений, представленных в таблице **Ошибка! Источник ссылки не найден.**.

Таблица 6.1.

Временные допустимые уровни ЭМП, создаваемые ПЭВМ

|  |  |  |
| --- | --- | --- |
| Наименование параметров | | ВДУ ЭМП |
| Напряженность | в диапазоне частот 5 Гц-2 кГц | 25 В/м |
| электрического поля | в диапазоне частот 2 кГц-400 кГц | 2,5 В/м |
| Плотность магнитного | в диапазоне частот 5 Гц-2 кГц | 250 нТл |
| потока | в диапазоне частот 2 кГц-400 кГц | 25 нТл |
| Электростатический потенциал экрана видеомонитора | | 500 В |

Мощность экспозиционной дозы мягкого рентгеновского излучения в любой точке на расстоянии 0,05 м от экрана и корпуса ВДТ (на электронно-лучевой трубке) при любых положениях регулировочных устройств не должна превышать одного мкЗв/час (100 мкР/час). В случае превышения данного показателя необходимо на ВДТ установить защитный экран.

СанПиН предъявляет ряд требований к визуальным параметрам, относящимся к их внешнему виду, дизайну, возможности настройки.

Для удобного считывания информации и подстройки под конкретного пользователя реализуется возможность настройки нужного положения монитора относительно взгляда наблюдателя путем регулирования его положения по горизонтали и вертикали. Данная возможность согласно СанПиН 2.2.2/2.4.1340-03 требуется как необходимая характеристика монитора. Кроме того, все современные ПК изготавливаются в спокойной цветовой гамме и покрываются специальным матовым покрытием для избежания солнечных бликов.

Для обеспечения комфортной работы пользователя определяются оптимальные и допустимые диапазоны визуальных эргономических параметров. Визуальные эргономические параметры ВДТ и пределы их изменений, в которых устанавливаются оптимальные и допустимые диапазоны значений, приведены в таблице 6.2.

Таблица 6.2.

Допустимые визуальные параметры устройств отображения информации.

|  |  |  |
| --- | --- | --- |
| **N** | **Параметры** | **Допустимые значения** |
| 1 | Яркость белого поля | Не менее 35 кд/м |
| 2 | Неравномерность яркости рабочего поля | Не более ± 20% |
| 3 | Контрастность (для монохромного режима) | Не менее 3:1 |
| 4 | Временная нестабильность изображения (непреднамеренное изменение во времени яркости изображения на экране дисплея) | Не должна фиксироваться |
| 5 | Пространственная нестабильность изображения (непреднамеренные изменения положения фрагментов изображения на экране) | Не более 2·10, где L - проектное расстояние наблюдения, мм |

Важнейшим визуальным эргономическим параметром, определяющим утомляемость зрительного аппарата, является мерцание изображения. СанПиН не регламентируют оптимальное значение частоты обновления, поскольку восприятие мерцания на экране монитора субъективно. Тем не менее, практика показывает, что в подавляющем большинстве случаев наиболее безопасна для здоровья частота обновления от 85 Гц и выше. Именно с учетом этой особенности проектируются все современные мониторы на ЭЛТ. Т.е. заявленному производителем «рабочему» разрешению монитора всегда соответствует частота обновления 85 Гц и выше (стандарты TCO’95 и TCO’99).

### 6.1.2. Требования к микроклимату, содержанию аэроинов и вредных химических веществ в воздухе на рабочих местах, оборудованных ПЭВМ

В помещениях, где работа с использованием ПЭВМ является основной, должны обеспечиваться параметры микроклимата на допустимом уровне в соответствии с СанПиН 2.2.4.548-96.

В помещениях, оборудованных ПЭВМ, проводится ежедневная влажная уборка и систематическое проветривание после каждого часа работы на ПЭВМ. Если параметры микроклимата не соответствуют установленным нормам, то необходимо применять системы кондиционирования. Работа за компьютером требует высокого сосредоточения и концентрации, поэтому более предпочтительными являются системы кондиционирования, автоматически поддерживающие заданный режим.

Уровни положительных и отрицательных аэроионов в воздухе помещений, где расположены ПЭВМ, должны соответствовать нормативам, установленным в СанПиН 2.2.4.1294-03.

Таблица 6.3. Показатели концентраций аэроинов.

|  |  |  |  |
| --- | --- | --- | --- |
| Нормируемые показатели | Концентрации аэроионов, ион/см3 | | Коэффициент униполярности |
| положительной полярности | отрицательной полярности |
| Минимально допустимые | p+ ≥ 400 | p->600 | 0,4≤ y < 1,0 |
| Максимально допустимые | p+<50000 | p- ≤ 50000 |

**6.1.3. Требования к уровням шума и вибрации на рабочих местах, оборудованных ПЭВМ.**

Допустимые уровни звукового давления отражены в таблице 6.4. Уровень звука, создаваемый ПЭВМ, не должен превышать 50 дБА.

Таблица 6.4.

Допустимые уровни звукового давления, создаваемого ПЭВМ.

|  |  |
| --- | --- |
| Среднегеометрическая частота октавной полосы, Гц | Допустимый уровень звукового давления, дБ |
| 31,5 | 86 |
| 63 | 71 |
| 125 | 61 |
| 250 | 54 |
| 500 | 49 |
| 1000 | 45 |
| 2000 | 42 |
| 4000 | 40 |
| 8000 | 38 |

Если использованы качественные комплектующие, то уровень звука будет значительно ниже. Наиболее шумными компонентами являются системный блок и оргтехника. С акустическими показателями оргтехники сложно что-то сделать, а для улучшения показателей системного блока можно использовать звукоизолирующие материалы либо корпус с более жесткой конструкцией, толстым материалом (один мм и более) и вентиляторами с меньшей скоростью вращения и большим диаметром.

При выполнении работ с использованием ПЭВМ в производственных помещениях уровень вибрации не должен превышать допустимых значений вибрации для рабочих мест (категория 3, тип «в») в соответствии с действующим санитарно-эпидемиологическим нормативом СН 2.2.4/2.1.8.566-96.

**6.1.4.Требования к освещению на рабочих местах, оборудованных ПЭВМ**

Помещения, оборудованные ПЭВМ и ВДТ, должны иметь естественное и искусственное освещение.

Рабочие столы следует размещать таким образом, чтобы видеодисплейные терминалы были ориентированы боковой стороной к световым проемам, чтобы естественный свет падал преимущественно слева. Светопроемы должны быть ориентированы преимущественно на север и северо-восток и обеспечивать коэффициент естественной освещенности (КЕО) не ниже 1.5% для наших широт.

Искусственное освещение в помещениях для эксплуатации ПЭВМ должно осуществляться системой общего равномерного освещения.

Освещенность на поверхности стола в зоне размещения рабочего документа должна быть 300 – 500 лк. Освещение не должно создавать бликов на поверхности экрана. Освещенность поверхности экрана не должна быть более 300 лк.

Следует также ограничивать прямую блесткость от источников освещения, при этом яркость светящихся поверхностей (окна, светильники и др.), находящихся в поле зрения, должна быть не более 200 кд/м2.

Следует ограничивать отраженную блесткость на рабочих поверхностях (экран, стол, клавиатура и др.) за чет правильного выбора типов светильников и расположения рабочих мест по отношению к источникам естественного и искусственного освещения, при этом яркость бликов на экране ПЭВМ не должна превышать 40 кд/м2 и яркость потолка не должна превышать 200 кд/м2.

Показатель ослепленности для источников общего искусственного освещения в производственных помещениях должен быть не более 20.

Яркость светильников общего освещения в зоне углов излучения от 50 до 90 градусов с вертикалью в продольной и поперечной плоскостях должна составлять не более 200 кд/м2, защитный угол светильников должен быть не менее 40 градусов.

Следует ограничивать неравномерность распределения яркости в поле зрения пользователя ПЭВМ, при этом соотношение яркости между рабочими поверхностями не должно превышать 3:1 – 5:1, а между рабочими поверхностями и поверхностями стен и оборудования 10:1.

В качестве источников света при искусственном освещении следует применять преимущественно люминесцентные лампы  типа ЛБ и компактные люминесцентные лампы. При устройстве отраженного освещения в производственных помещениях допускается применение металлогалогенных ламп.

Для освещения помещений с ПЭВМ следует применять светильники с зеркальными параболическими решетками, укомплектованными электронными пуско-регулирующими аппаратами (ЭПРА). Допускается использование многоламповых светильников с электромагнитными пуско-регулирующими аппаратами (ЭПРА), состоящими из равного числа опережающих и отстающих ветвей. Применение светильников без рассеивателей и экранирующих решеток не допускается.

При отсутствии светильников с ЭПРА лампы многоламповых светильников или рядом расположенные светильники общего освещения следует включать на разные фазы трехфазной сети.

Общее освещение при использовании люминесцентных светильников следует выполнять в виде сплошных или прерывистых линий светильников, расположенных локализовано над рабочим столом ближе к его переднему краю, обращенному к оператору, параллельно линии зрения пользователя при периметральном расположении видеодисплейных терминалов.

Коэффициент запаса (Кз) для осветительных установок общего освещения должен приниматься равным 1,4.

Коэффициент пульсации не должен превышать 5%.

Для обеспечения нормируемых значений освещенности в помещениях для использования ПЭВМ следует проводить чистку стекол оконных рам и светильников не реже двух раз в год и проводить своевременную замену перегоревших ламп.

**6.1.5. Требования к организации и оборудованию рабочих мест с ПЭВМ**

Экран видеомонитора должен находиться от глаз пользователя на расстоянии 600-700 мм, но не ближе 500 мм с учетом размеров алфавитно-цифровых знаков и символов.

Высота рабочей поверхности стола для взрослых пользователей должна регулироваться в пределах 680-800 мм, при отсутствии такой возможности высота рабочей поверхности стола должна составлять 725 мм.

Модульными размерами рабочей поверхности стола для ПЭВМ, на основании которых должны рассчитываться конструктивные размеры, следует считать: ширину 800, 1000, 1200 и 1400 мм; глубину 800 и 1000 мм при нерегулируемой его высоте, равной 725 мм.

Рабочий стол должен иметь пространство для ног высотой не менее 600 мм, шириной – не менее 500 мм, глубиной на уровне колен – не менее 450 мм и на уровне вытянутых ног – не менее 650 мм.

Конструкция рабочего стула должна обеспечивать:

* ширину и глубину поверхности сиденья не менее 400 мм;
* поверхность сиденья с закругленным передним краем;
* регулировку высоты поверхности сиденья в пределах 400 – 550 мм и углам наклона вперед до 15 град., и назад до 5 град.;
* высоту опорной поверхности спинки 300 ± 20 мм, ширину – не менее 380 мм и радиус кривизны горизонтальной плоскости – 400 мм;
* угол наклона спинки в вертикальной плоскости в пределах ± 30 градусов;
* регулировку расстояния спинки от переднего края сиденья в пределах 260 – 400 мм;
* стационарные или съемные подлокотники длиной не менее 250 мм и шириной 50 – 70 мм;
* регулировка подлокотников по высоте над сиденьем в пределах 230 +- 30 мм и внутреннего расстояния между подлокотниками в пределах 350 – 500 мм.

Поверхность сиденья, спинки и других элементов стула (кресла) должна быть полумягкой, с нескользящим, слабо электризующимся и воздухопроницаемым покрытием, обеспечивающим легкую очистку от загрязнений.

Клавиатуру следует располагать на поверхности стола на расстоянии 100 – 300 мм от края, обращенного  к пользователю или на специальной,  регулируемой по высоте рабочей поверхности, отделенной от основной столешницы.

**6.1.6. Пожарная безопасность**

Согласно существующей классификации ОНТП-24-86 помещение, оборудованное ПЭВМ, по пожарной опасности относится к категории «Д». Пожарная безопасность обеспечивается мерами пожарной профилактики и активной пожарной защитой. Понятие о пожарной профилактике включает комплекс мероприятий, необходимых для предупреждения возникновения пожара или уменьшения его последствий. Под активной пожарной защитой понимаются меры, обеспечивающие успешную борьбу с возникающими пожарами.

Для эффективного тушения пожаров в помещении, оборудованном ПЭВМ, необходимо применять углекислотные и порошковые огнетушители, которые обладают возможностью тушения электроустановок и высокой эффективностью борьбы с огнем.

В качестве пожарных сигнальных датчиков в помещении устанавливаются дымовые фотоэлектрические извещатели типа ПОСТ-1. Исходя из высоты потолка (2.6 м) и площади помещения (18 м2) по нормам достаточно двух извещателей.

**6.1.7. Электробезопасность**

Поражения электрическим током является наиболее опасным потенциальным фактором при работе с электрооборудованием.

К опасным производственным факторам относится напряжение в электрической сети (ГОСТ 12.1.038-82). Проходя через организм, электрический ток оказывает термическое, электролитическое и биологическое действие. В результате могут возникнуть различные нарушения и даже полное прекращение деятельности органов дыхания и кровообращения.

Для обеспечения электробезопасности применяют заземление электрических устройств и устройства защитного отключения.

Напряжение в сети составляет 220В, а частота переменного тока 50Гц, и согласно ПУЭ помещение относится ко 2-ой категории, т.е. к помещению с повышенной опасностью.

Исходя из этого, для обеспечения безопасности предполагается использование заземления. Заземлению подлежат корпуса компьютеров и периферийных устройств.

В основе действия устройств защитного отключения, как электрозащитного средства, лежит принцип ограничения (за счет быстрого отключения) продолжительности протекания тока через тело человека при непреднамеренном прикосновении его к элементам электроустановки, находящимся под напряжением. Из всех известных электрозащитных средств устройства защитного отключения являются единственными, обеспечивающими защиту человека от поражения электрическим током при прямом прикосновении к одной из токоведущих частей. В условиях, в которых выполнялся проект устройство защитного отключения можно было использовать как дополнительную защиту.

### 6.2. Расчет освещенности на рабочем месте

Задачей данного светотехнического расчета является определение мощности всей осветительной установки для получения заданной освещенности на рабочем месте, при выбранном типе и расположении светильников.

**Исходные данные:**

*Источник света:* лампы люминесцентные;

*Система освещенности:* общая равномерного освещения;

*Общий уровень освещенности:* 500 лк;

*Размеры помещения:* ширина 3 м, длина 6 м, высота 2.6 м

площадь S=18 м2, объем V=46.8 м3

**Определить:** Тип и число ламп осветительной системы;

общую мощность осветительной системы;

**Расчет:**

Расчет следует проводить по методу коэффициента использования светового потока. Необходимый световой поток лампы в каждом светильнике, согласно [6]:

*Fо* = , где

*Fо* - рассчитываемый световой поток лампы, лм;

*Е* - заданная минимальная освещённость, лк (500);

N - число светильников;

- коэффициент запаса, учитывающий уменьшение светового потока лампы в результате загрязнения светильников в процессе эксплуатации (для люмин. ламп - 1,4) ;

*S* - освещаемая площадь, м2 (S=18);

*z* - отношение средней освещённости к минимальной (для люмин. ламп - 1,1) ;

*η*  - коэффициент использования светового потока в долях единицы (отношение светового потока, падающего на расчётную поверхность, к суммарному потоку всех ламп).

Коэффициент использования *η* зависит от типа светильника, от коэффициентов отражения потолка ρп, стен ρс, расчётной поверхности ρр, индекса помещения:

*i*  = , где

*h* - высота светильника над рабочей поверхностью,

*а* - длина помещения, *b* - ширина помещения.

*h* = H-hp-hc*,* , где

H = 2.5 *-* высота помещения, м;

hp *=* 0.7 - высота рабочей поверхности от пола, м;

hc*=* 0.3 - высота свеса светильника от основного потолка, м.

Таким образом, имеем:

*h = H-hp-hc = 4-0.7-0.3 = 1.5 (м)*

*i =*= *= 1.33 -> 1.5*

Для светлого фона стен и потолка примем: *ρп = 70, ρс = 50;*

Следовательно, по таблице для *i* =1.5 находим *η=57%*

Предварительно возьмем количество ламп *N* = 6 , а если в результате расчета не удастся подобрать необходимый тип ламп, придется его подкорректировать.

*Fо* =  *= 4052 лм;*

Для обеспечения светового потока в 4052 лм подходит люминесцентная лампа марки ЛД80 (мощность 80 Вт, световой поток 4070 лм).

Отклонение потока выбранной лампы от расчетного составляет:

*,* что лежит в пределах нормы от -10% до 20%. Следовательно, марка лампы подобрана верно.

Число светильников выбирается в зависимости от размеров освещаемого помещения, при этом количество светильников должно быть таким, чтобы отношение расстояния между ними к высоте их подвеса над поверхностью было равно 1,5 ÷ 2 . При выборе осветительных приборов используем светильники типа ЛСП44 (длина светильника 1580 мм, ширина 97 мм). Каждый светильник комплектуется двумя лампами. Размеры помещения позволяют разместить светильники в один ряд. Общее число ламп в светильниках N=6.

Для исключения засветки экранов дисплеев прямыми световыми потоками светильники общего освещения располагают сбоку от рабочего места, параллельно линии зрения и оператора стене с окнами. Такое размещение светильников позволяет производить их последовательное включение в зависимости от величины естественной освещённости и исключает раздражение глаз чередующимися полосами света и тени, возникающее при поперечном расположении светильников.

Электрическая мощность всей осветительной системы вычисляется по формуле:

, Вт, где

*P1* = 80 Вт – мощность одной лампы;

*N* = 6 – количество ламп.

Вт

**Результат:** Требуемое количество ламп – 6, мощность осветительной системы 480 Вт.

### 6.3. Расчет снижения количества выхлопных газов автомобиля

При использовании алгоритма прогнозирования пробок в автомобильном навигаторе ожидается снижение времени пути автомобиля.

Рассмотрим, насколько снизится количество выхлопных газов, выделяемых в автомобилях в Москве при условии, что все автомобилисты будут использовать оптимальный маршрут пути и меньше времени затрачивать на поездку. Для этого используем API Яндекс.Карт [7], позволяющий строить маршруты в Москве с учетом пробок и без учета.

Выберем случайным образом 500 пар точек в Москве и Московской области, и для каждой пары проложим два вида рассматриваемых маршрутов.

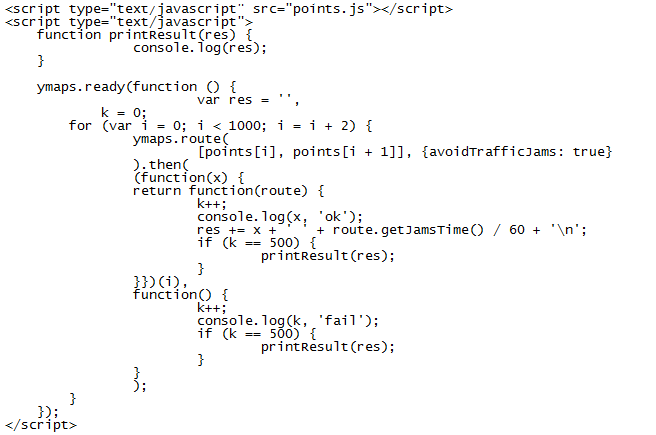


Рис. 6.1. Фрагмент программы на JavaScript, осуществляющий запрос за данными.

Были проведены испытания в выходные и будние дни.

Результаты выходного дня не очень показательные, ввиду того, что на дороге практически нет затруднений. Тем не менее, применение алгоритма построения маршрутов с учетом пробок дал выигрыш по времени в **1%** по сравнению с алгоритмом прокладки маршрутов без учета пробок.

Рассмотрим полученные результаты для буднего дня.

Рис. 6.2. Сравнение времени в пути следования по маршрутам

На графике заметно, что практически во всех случаях наблюдается выигрыш по времени при прокладке маршрута с учетом пробок.

Абсолютный средний выигрыш по времени пути составил 3.74 мин.

Относительный средний выигрыш по времени пути составил 3.87%.

Предполагается, что прогнозирование пробок улучшит эти показатели при внедрении системы в устройства автомобильной навигации.

Рассчитаем процент снижения времени пути за неделю с учетом полученных результатов.

Процент снижения времени пути в выходной день составил 1%.

Процент снижения времени пути в будний день составил 3.87%.

Примем, что в неделе 5 рабочих дней и 2 выходных дня (не будем учитывать праздничные дни).

Недельный выигрыш по времени пути составит в таком случае

(1 \* 2 + 3.87 \* 5) / (2 + 5) = 3.05% ≈ 3%

По данным исследования за 1998 год в атмосферу Москвы ежедневно попадало около 4.5т загрязняющих веществ. [8].

Примем, что количество загрязняющих веществ, выброшенных автомобилем, линейно зависит от времени поездки.

При среднем снижении времени пути автомобиля на 3%, получим следующие результаты [8].

Таблица 6.5.

Реальное и прогнозируемое количество вредных выбросов.

|  |  |  |
| --- | --- | --- |
| Загрязняющее вещество | Объем выбросов, т/сут | Прогнозируемый объем выбросов, т/сут |
| CO | 3607.5 | 3499,275 |
| углеводороды | 744.1 | 721,777 |
| Оксиды азота | 281.4 | 272,958 |
| Оксиды серы | 27.5 | 26,675 |
| Свинец | 0.155 | 0,15035 |
| Сажа | 45.6 | 44,232 |
| Всего | 4704.155 | 4563,03 |

Рис 6.3. Гистограмма снижения уровня выхлопных газов при оптимизации маршрутов.

Данные за 1998г. можно считать опорными для определения выгоды от использования оптимизированного алгоритма маршрутизации в будущем.

При оптимизации можно получить снижение количества выбрасываемых загрязняющих веществ на несколько тонн в сутки.

# Заключение

В результате выполнения данной дипломной работы была получена математическая модель, позволяющая осуществлять предсказания скорости на основе исторических данных на час вперед с интервалами между предсказаниями 15 минут.

Была решена проблема неполноты данных.

Разработанная модель предоставляет прогноз, превосходящий по точности алгоритм прогнозирования средним значением на ребрах, для которых плотность исторических данных превышает 0.1 (известен хотя бы каждый 10й отсчет). Средняя величина ошибки прогнозирования составила ≈ 10км/ч.

# Литература

1. ГОСТ 12.1.038-82 «Электробезопасность»
2. Гуда С., Рябов Д., «Прогнозирование пробок на улицах по известным данным о скорости автомобилей», Южный Федеральный Университет
3. Документация к языку Python 3 [электронный ресурс]. Режим доступа: python.org, свободный.
4. Изучение уровня загрязненности воздуха в Москве [электронный ресурс]. Режим доступа: <http://student.km.ru/ref_show_frame.asp?id=346F7183413544D1B4560C38406C0506>, свободный.
5. **Кнорринг Г. М. и др.** Справочная книга для проектирования электрического освещения / Г. М. Кнорринг, И. М. Фадин, В. Н. Сидоров — 2-е изд., перераб. и доп. — СПб.: Энергоатомиздат. Санкт-Петербургское отд-ние, 1992. —448 с: ил.
6. Прохоров С.А. «Прикладной анализ неэквидистантных временных рядов». Самарский государственный аэрокосмический университет, 2001.
7. Пупырев С., Пронченков А., «Прогнозирование загруженности автомобильных дорог», Уральский государственный университет, Екатеринбург
8. СанПиН 2.2.4.1294-03 «Гигиенические требования к аэроионному составу воздуха производственных и общественных помещений»
9. СанПиН 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений
10. СанПин 2.2.2/2.4.1340-03. **"Гигиенические требования к персональным электронно-вычислительным машинам и организации работы»**
11. СН 2.2.4/2.1.8.566-96. Производственная вибрация. Вибрации в жилых и общественных зданиях. Санитарные нормы.
12. API Яндекс.Карт [электронный ресурс]. Режим доступа: api.yandex.ru/maps, свободный.
13. S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. In Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), San Diego, CA, 1999.
14. H. Gonzalez, J. Han, X. Li, M. Myslinska, and J. P. Sondag. Adaptive fastest path computation on a road network: A traffic mining approach. In Proceedings of the 33nd International Conference on Very Large Data Bases (VLDB), Vienna, Austria, 2007.9. E. Horvitz, J. Apacible, R. Sarin and L. Liao. Prediction, Expectation, and Surprise: Methods, Designs, and Study of a Deployed Traffic Forecasting Service. Microsoft Research University of Washington. Redmond, Washington
15. M. Hadjieleftheriou, G. Kollios, D. Gunopulos, and V. Tsotras. On-line discovery of dense areas in spatiotemporal databases. 2003.
16. J. Lee, J. Han, and K. Whang. Trajectory clustering: A partition-and-group framework. In Proceedings of the SIGMOD Conference, Beijing, China, 2007.
17. X. Li, J. Han, J.-G. Lee, and H. Gonzalez. Traffic density-based discovery of hot routes in road networks. In Proceedings of the 10th International Symposium on Spatial and Temporal Databases (SSTD), Boston, MA, 2007.
18. P. Kalnis, N. Mamoulis, and S. Bakiras. On discovering moving clusters in spatio-temporal data. pages 364–381, 2005.
19. R. K. Oswald, W. T. Scherer, and B. L. Smith. Traffic flow forecasting using approximate nearest neighbor nonparametric regression. In Final project of ITS Center project: Traffic forecasting: non-parametric regressions, December, 2000.
20. S. Shekhar, C.-T. Lu, S. Chawla, and P. Zhang. Data mining and visualization of twin-cities traffic data. In Technical Report TR 01-015, Dept. of CSE, Univ. of Minnesota, 2000.