

В.В. Неприенков, Р.А. Сакаев, А.А. Ступников

Тюменский государственный университет, г. Тюмень
УДК 004.9

СОЗДАНИЕ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ОБНАРУЖЕНИЯ НЕПОЛАДОК НА СЕТЯХ ВОДОСНАБЖЕНИЯ Г. ТЮМЕНИ

Аннотация. В статье описан процесс разработки клиент-серверного приложения для обнаружения нештатных ситуаций на водопроводных сетях на основе анализа показаний датчиков воды и оповещения о возникших авариях пользователей системы.

Ключевые слова: машинное обучение, модели классификации, водоснабжение, мониторинг, клиент-серверное приложение, обработка данных.

Введение

Для поставки качественных услуг по водоснабжению, необходимо контролировать каждый участок водопроводной сети на предмет нештатных ситуаций. Однако это является крайне трудной задачей для человека из-за большого количества расположенных на сетях датчиков воды, которые могут измерять и передавать различные показатели.

На сегодняшний день все нежелательные ситуации на водопроводных сетях регулируются путем выезда бригады на место аварии при ее непосредственном обнаружении, либо при обнаружении начальником центральной диспетчерской службы показателей датчиков, которые отклоняются от нормы.

Существующий способ обнаружения и определения аварийных ситуаций на сетях водоснабжения отнимает большое количество временных и трудовых ресурсов, и в некоторых случаях реагирование на

ситуацию может быть несвоевременным, что может привести к потере водных ресурсов [1]. Учитывая тот факт, что показания датчиков собираются автоматически и регулярно, то выявление нештатных ситуаций может быть реализовано на уровне автоматизированной системы.

Идея решения

В качестве решения поставленной проблемы предлагается разработать клиент-серверное приложение для обнаружения нештатных ситуаций на водопроводных сетях и оповещения о них пользователей системы.

Серверная часть приложения в автоматическом режиме будет опрашивать базу данных на предмет новых показаний счетчиков. При получении новых показаний система на основе исторических данных должна определить, являются ли переданные показания приемлемыми. В случае, если новые данные отклоняются от приемлемых значений, необходимо определить вид нештатной ситуации и оповестить о ней пользователя системы.

Клиентская часть помимо оповещения о нештатных ситуациях, должна отображать приборы и сети водоснабжения на карте города и предоставлять информацию о работе выделенного участка сети в наглядном и понятном пользователю виде. Предоставленная информация должна быть исчерпывающей, чтобы пользователь мог принять оперативное решение по устранению неполадки или проверки подозрительного участка сети.

Средства разработки

В качестве СУБД используется MS SQL SERVER, так как именно данное программное обеспечение используется на предприятии «Тюмень Водоканал» для хранения показаний приборов учета. В качестве

серверного языка и языка для создания моделей классификации был использован Python. Клиентская часть приложения написана на языке HTML в связке с CSS и Javascript. В качестве поставщика географических данных был выбран OpenStreetMaps (OSM). По сравнению с Яндекс API, данные которого нельзя использовать для закрытых проектов, где требуется регистрация, и Google API, использование данных которого может закончиться финансовыми расходами, если не ограничивать число запросов, использовать OSM можно для любых нужд, при этом не беспокоясь о финансовой стороне использования географических данных.

Построение моделей классификации

Первый шаг — это получение и подготовка выборки данных для дальнейшей разработки. Первоначальные данные содержат показания счетчиков воды. Выборка состоит из множества показателей, многие из которых являются неинформативными для определения нештатных ситуаций. Важными для задачи классификации являются следующие данные: «Метка» (идентификатор прибора), «ДатаВремя» (дата и время передачи показания), «V3» (объем проходящей воды), «Т3» (температура воды), «Р3» (давление), «Class» (вид нештатной ситуации).

Очистка данных подразумевает работу с шумами в выборке. В данном случае все шумы внутри выборки означают нештатную ситуацию, вне зависимости от того какое количество времени данные не являлись нормальными.

Строки, заполненные NULL-значениями, не могут быть обработаны моделью, так как NULL не является числом, поэтому для анализа и построения классификаторов эти значения будут заменены нулями. При подобном преобразовании значения будут корректно восприняты моделью.

После очистки выборки ее необходимо преобразовать. Данный этап подготовки данных заключается в нормализации числовых значений внутри полей, дискретизации значений, а также выделении новых признаков [4]. Нормализация каждого числового значения происходит по формуле:

$$X_{i \text{ норм}} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}$$

После такого преобразования все числовые значения будут безразмерными и находиться в диапазоне от 0 до 1, где 0 – минимальное значение признака, 1 – максимальное. В качестве нового признака было выделено время суток передачи показания. Необходимость выделения данного признака состоит в том, что нормы показателей для разных приборов в дневное и ночное время могут кардинально различаться [2]. Для улучшения работы моделей классификации была проведена дискретизация некоторых полей. Дискретизация – процесс, при котором все числовые значения в столбце выборки будут разбиты на группы, а их реальные значения заменены на номер группы.

В таблице 1 представлены результаты тестирования разных моделей классификации [5]. Размер тестовой выборки составлял 65% от всего числа записей. Выбранной метрикой для оценки работы классификаторов является ассурасу (количество верно классифицированных объектов относительно общего количества всех объектов). В таблице приведен процент точности определения класса показаний для чистой выборки, то есть набора данных, над которым не было произведено операций преобразования, а также для выборки с применением нормализации, дискретизации и выделения нового признака.

Таблица 1. Сравнение результатов классификаторов.

Метод	Чистая выборка	Все методы
Случайный лес	72.36	84.35
Дерево решений	71.50	83.70
К-ближайших соседей	73.23	86.51
Логистическая регрессия	71.37	77.35
Градиентный спуск	74.33	85.37

Классификатор, использующий метод k-ближайших соседей, показал самые лучшие результаты. Поэтому модели этого классификатора были обучены для каждого отдельного прибора и экспортированы в файлы для их дальнейшего использования в приложении.

Механизм классификации поступающих показаний

Суть механизма, который классифицирует поступающие показания датчиков состоит в следующем. Сервер приложения в фоновом режиме периодически опрашивает базу данных на наличие новых показаний, которые можно было бы передать в классификатор. Для этого на сервере хранится максимальный идентификатор среди записей таблицы показаний приборов на момент запуска сервера. Затем с определенной периодичностью отправляется запрос в базу для получения текущего максимального идентификатора. В случае, если значение идентификатора в базе больше, чем на сервере, то это означает появление в базе новых показаний, которые необходимо передать в классификатор. Для каждой новой записи из файла импортируется обученная модель классификатора. Полученный результат, то есть класс показания, записывается в базу данных.

Отображение элементов водопроводной сети на карте

Для отображения приборов учета на карте сперва было необходимо получить их координаты при помощи геокодера. Для рисования сетей водоснабжения был подключен плагин, который позволяет добавлять геометрические объекты на карту и управлять ими. Предприятие “Тюмень Водоканал” предоставило схему расположения труб интересующих районов, с помощью которой на карте вручную были отрисованы трубы.

Стоит отметить, что большое количество одновременно отображаемых на карте объектов нагружает клиентское приложение, поэтому для решения данной проблемы маркеры были объединены в кластеры на основе расстояний между ними. Числа на маркерах показывают число дочерних элементов внутри кластера (Рисунок 1).

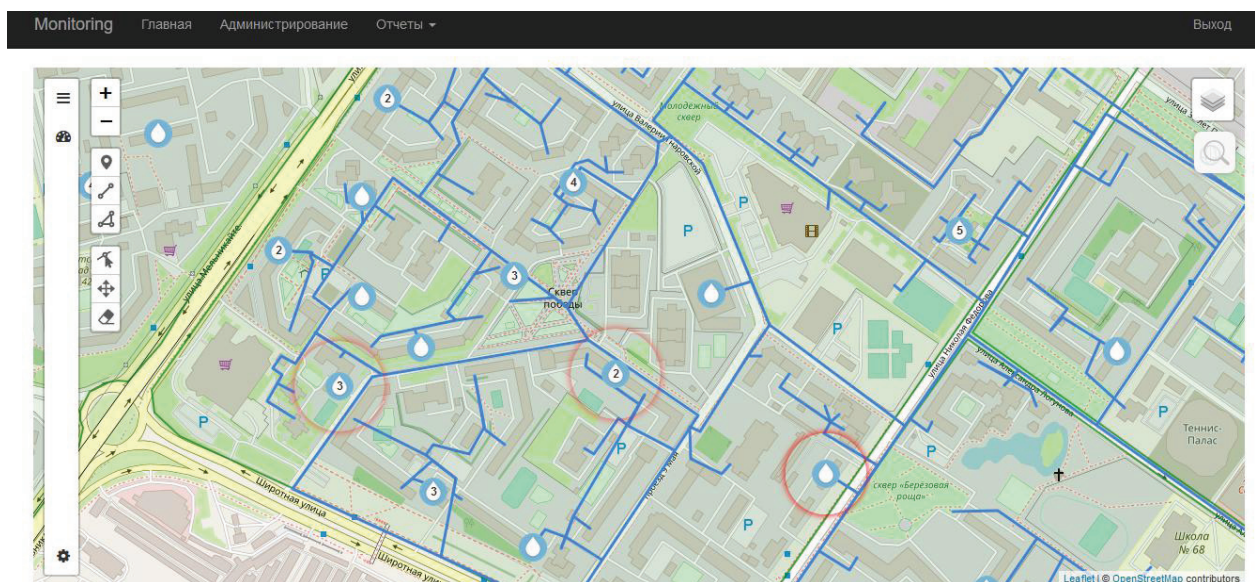


Рис. 1. Карта с объединенными в кластеры маркерами.

Визуализация показаний прибора

Для полноценного мониторинга сетей водоснабжения необходимо отображать пользователю показания приборов в разных форматах. При нажатии на маркер открывается панель, содержащая информацию о работе прибора. Используется два вида представления данных: в виде диаграммы, а также в табличном виде (Рисунок 2). Для обоих видов есть возможность фильтрации периода, за который отображаются данные.

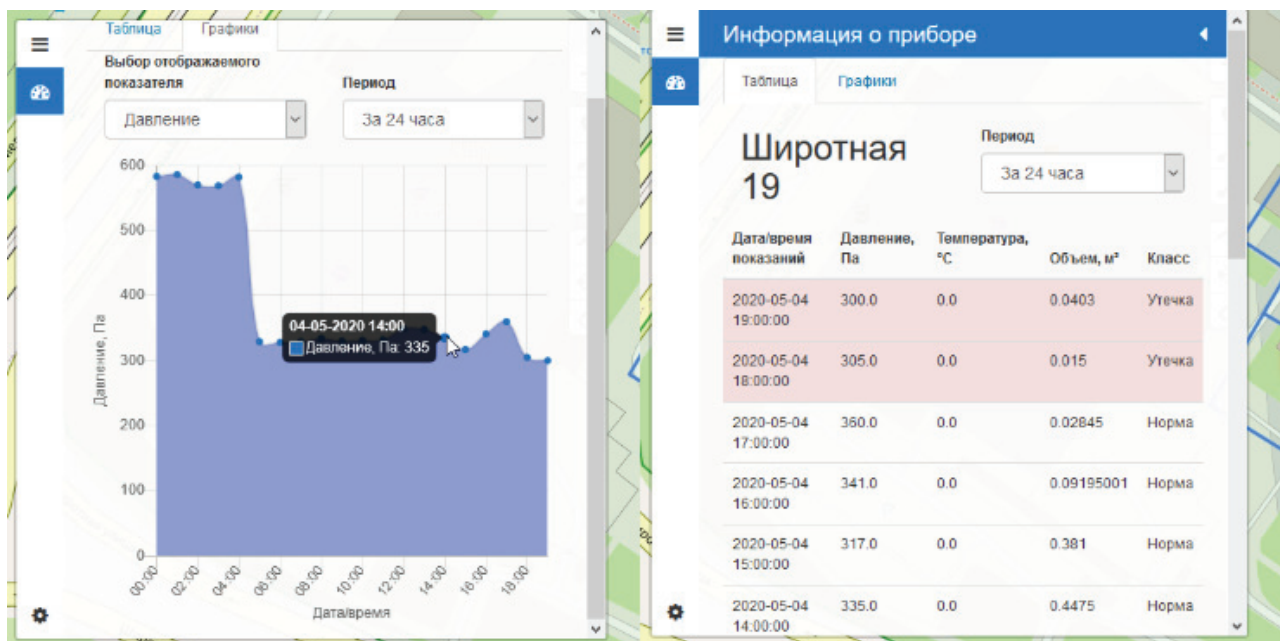


Рис. 2. Визуализация показаний прибора.

Если у пользователя открыта данная вкладка, то на сервер периодически будет поступать запрос, который проверяет наличие новых показаний в базе данных, которые можно было бы отобразить в таблице или на графике. Красным цветом выделены те показания, которые классификатор посчитал отличающимися от нормы.

Оповещение пользователей о нештатных ситуациях

У пользователя должно быть несколько способов узнать о возникшей нештатной ситуации. Одним из способов является отображение списка аварий в табличном виде. Список аварий можно получить, открыв боковую панель (Рисунок 3). В таблице отображается время передачи показания, адрес прибора, а также тип нештатной ситуации, который присвоил классификатор последнему переданному показанию прибора. Если нештатная ситуация новая, то она помечается красным цветом. Аварии отсортированы в таблице таким образом, что сначала отображаются новые аварии, отсортированные по времени.

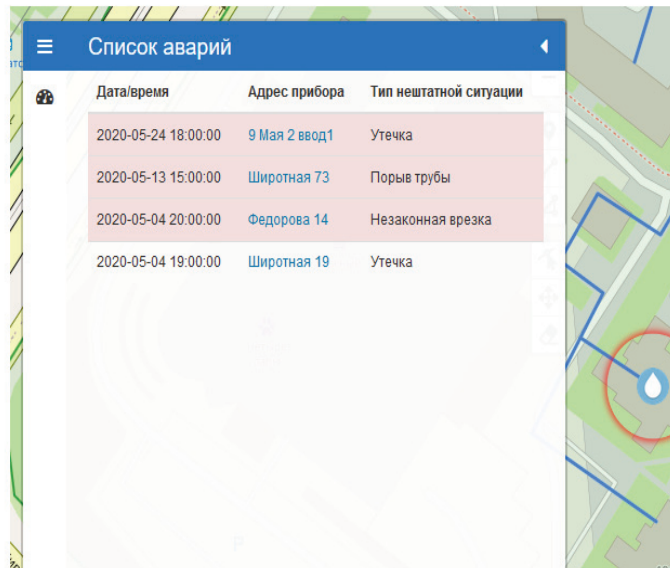


Рис. 3. Список аварий.

Информацию о новых авариях клиентское приложение получает путем периодических аякс-запросов. Если таблица с авариями закрыта, то оповещения о нештатных ситуациях появляются в виде всплывающих сообщений в правой нижней части экрана (Рисунок 4). Всплывающее сообщение содержит в себе дату, когда нештатные ситуации на сети были обнаружены, а также количество новых аварий на указанный момент времени.

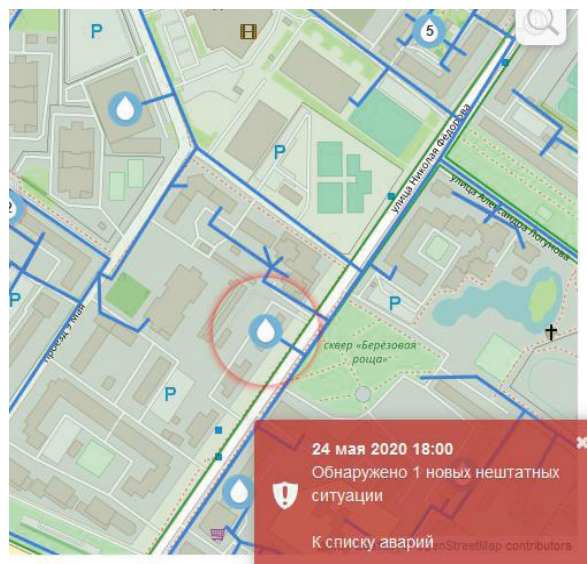


Рис. 4. Всплывающее сообщение об обнаруженной нештатной ситуации.

Приборы, чьи последние показания, были классифицированы как аварийные по-особому отображаются на карте. При этом особый вид отображения распространяется на иконку всего кластера. Поэтому в случае, если хотя бы один дочерний прибор внутри кластера будет передавать аварийные показания, то вся иконка будет пульсировать, что позволит находить аварийные приборы при любом масштабе карты. Если новое показание прибора будет классифицировано как неаварийное, то пульсация пропадает.

Заключение

Использование разработанного приложения позволяет сократить временные, трудовые, а также водные ресурсы за счет уменьшения времени реагирования на возникающие нештатные ситуации благодаря высокой точности определения нештатных ситуаций на основе поступающих показаний приборов.

СПИСОК ЛИТЕРАТУРЫ

1. Громов Г. Н., Примин О. Г., Бычков Д. А. Модель расчета потерь и неучтенных расходов воды в водопроводной сети г. Тюмени // Водоснабжение и санитарная техника. 2016. №9. С.16-22.
2. СНиП 2.04.02-84 Водоснабжение. Наружные сети и сооружения. Госстрой СССР, Москва, 1985. – 136 с.
3. СП 31.13330.2012 Водоснабжение. Наружные сети и сооружения. Актуализированная редакция СНиП 2.04.02-84 (с Изменениями N 1, 2)
4. Митин И. В., Русаков В. С. Анализ и обработка экспериментальных данных- 5-е издание. М.: Форум, 2012. – 24с.

5. Scikit-learn Classification [Электронный ресурс]: официальный сайт библиотеки scikit-learn/ раздел классификации данных. URL: <https://scikit-learn.org> (Дата обращения 10.05.2020).