

ОЧИСТКА СЕНСОРНЫХ ДАННЫХ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ УПРАВЛЕНИЯ ОТОПЛЕНИЕМ ЗДАНИЙ

© 2021 М.Л. Цымблер, Я.А. Краева, Е.А. Латыпова,

Е.В. Иванова, Д.А. Шнайдер, А.А. Басалаев

Южно-Уральский государственный университет

(454080 Челябинск, пр. им. В.И. Ленина, д. 76)

E-mail: mzym@susu.ru, kraevaya@susu.ru, latypovaea@susu.ru,

elena.ivanova@susu.ru, shnaiderda@susu.ru, basalaeava@susu.ru

Поступила в редакцию: 03.09.2020

В современных интеллектуальных системах управления отоплением зданий зачастую возникают пропуски значений или выбросы в показаниях температурных и других датчиков ввиду сбоев программного или аппаратного обеспечения либо человеческого фактора. Для обеспечения эффективного анализа данных и принятия решений некорректные данные датчиков следует очищать путем восстановления пропущенных значений и сглаживания выбросов. В данной статье представлен пример SCADA-системы ПолиТЭР для управления отоплением, установленной в Южно-Уральском государственном университете, и описана структура и принципы реализации Модуля очистки данных, внедренного в указанную систему. Модуль очистки данных реализован с помощью технологий интеллектуального анализа данных и нейронных сетей в виде набора следующих подсистем. Препроцессор извлекает необработанные данные из хранилища данных системы и подготавливает обучающий набор данных для дальнейшей обработки. Предиктор представляет собой рекуррентную нейронную сеть для прогнозирования следующего значения датчика на основе его исторических данных. Реконструктор определяет, является ли текущее значение датчика выбросом, и в таком случае заменяет его на синтетическое значение, полученное Предиктором. Наконец, Детектор аномалий в режиме реального времени обнаруживает аномальные промежутки в данных датчика. В вычислительных экспериментах на реальных данных разработанный модуль показал относительно высокую и стабильную точность, а также адекватное обнаружение аномалий.

Ключевые слова: умный дом, температурный датчик, управление отоплением, хранилище данных, очистка данных, временной ряд, восстановление пропущенных значений, поиск выбросов, обнаружение аномалий, рекуррентная нейронная сеть.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Цымблер М.Л., Краева Я.А., Латыпова Е.А., Иванова Е.В., Шнайдер Д.А., Басалаев А.А. Очистка сенсорных данных в интеллектуальных системах управления отоплением зданий // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2021. Т. 10, № 3. С. 16–36. DOI: 10.14529/cmse210302.

Введение

Современные интеллектуальные системы управления отоплением зданий используют различные источники данных, включая данные измерений счетчиков потребления коммунальных услуг, данные от контроллеров процесса, датчиков внутреннего климата и др. Внедрение технологий Интернета вещей (IoT, Internet of Things) в системы такого типа дополнительно позволяет получать большие массивы данных о различных параметрах, влияющих на общие условия отопления в здании. Полученный в результате массив больших данных позволяет проводить всесторонний анализ систем отопления, позволяя своевременно выявлять случаи отклоняющихся значений производительности и показателей энергоэффективности зданий за допустимые пределы. Однако, в соответствии с теорией

надежности при рассмотрении всей совокупности данных в единой системе увеличение количества источников данных может привести к выполнению некорректного анализа данных в случае сбоя даже одного источника данных.

Причиной возникновения некорректных показаний IoT-датчиков могут являться различные факторы: износ оборудования, неправильные монтаж и эксплуатация оборудования (человеческий фактор), конструктивные недостатки программного и аппаратного обеспечения, выход условий эксплуатации за пределы допустимых значений и др. Некорректные показания IoT-датчиков являются причиной снижения энергоэффективности эксплуатации объектов и их некорректного управления в автоматическом и автоматизированном режиме, а также приводят к штрафным санкциям от поставщиков энергоносителей. В соответствии с этим актуальной задачей является разработка методов и алгоритмов очистки данных, обеспечивающих оперативное выявление некорректных показаний датчиков и восстановление утраченных данных.

В настоящей статье рассматривается проблема очистки данных в интеллектуальной системе управления теплоснабжением зданий кампуса Южно-Уральского государственного университета (ЮУрГУ). В кампусе ЮУрГУ реализована интеллектуальная система управления теплоснабжением зданий на основе SCADA-системы ПолиТЭР [1], которая позволяет выполнять мониторинг инженерных систем зданий кампуса и управление режимами их работы, включая проводные и беспроводные IoT-датчики. Авторами статьи разработан и внедрен в систему ПолиТЭР Модуль очистки данных, обеспечивающий в режиме реального времени выявление некорректных показаний и восстановление утраченных данных IoT-датчиков.

Статья организована следующим образом. В разделе 1 приведен обзор SCADA-системы «ПолиТЭР». В разделе 2 приведен краткий обзор научных работ в данной области. Раздел 3 посвящен описанию архитектуры и принципов реализации Модуля очистки данных для SCADA-системы «ПолиТЭР». Результаты вычислительных экспериментов, исследующих эффективность разработанного модуля, представлены в разделе 4. В заключении подводятся итоги проведенного исследования.

1. Обзор системы ПолиТЭР

На рис. 1 представлена структура интеллектуальной системы управления теплоснабжением зданий кампуса ЮУрГУ. На нижнем уровне система включает в себя различные проводные и беспроводные датчики, приборы учета и контроллеры. На среднем уровне выполняется обеспечение связи контроллеров и узлов учета с сервером баз данных посредством различного проводного и беспроводного сетевого оборудования. Третий уровень включает SCADA-систему «ПолиТЭР» с сервером баз данных, осуществляющую обработку информации. Представление обработанных данных осуществляется на АРМ локальных пользователей через ЛВС ЮУрГУ или на АРМ удаленных пользователей через сеть Интернет.

Внедрение данной системы в ЮУрГУ началось в 2010 г. Особое внимание при внедрении данной системы в ЮУрГУ было уделено системе теплоснабжения. Целью внедрения стала комплексная оптимизация процессов управления теплоснабжением и теплопотреблением за счет интеграции территориально распределенного измерительного оборудования с автоматическими системами управления собственными системами когенерации и распределительными системами как централизованно, так и индивидуально на отдельных потребителях.

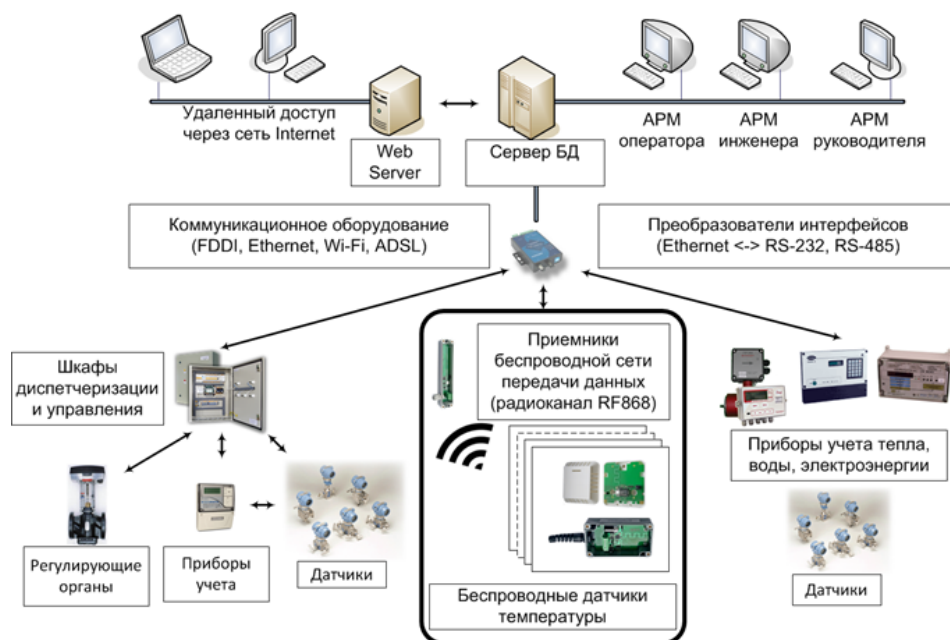


Рис. 1. Архитектура SCADA-системы ПолиТЭР

Интеллектуальной системы управления теплоснабжением зданий включают в себя следующее измерительное и управляющее оборудование:

- узлы учета тепловой энергии (57 шт.), оборудованные тепловычислителем с подключенными к нему 2 датчиками давления, 2 датчиками температуры и 2 расходомерами;
- узлы учета холодной воды (58 шт.), оборудованные вычислителем с подключенными к нему датчиком давления и расходомером;
- узлы учета газа (3 шт.), оборудованные газовым корректором с подключенными к нему расходомером датчиками температуры и давления;
- контроллеры управления распределением тепловой энергии в индивидуальных тепловых пунктах (24 шт.) с подключенными к ним 5 датчиками температуры и различными дискретными датчиками защит, а также управляющими приводами.
- контроллеры управления генерацией и распределением тепловой и электрической энергии (5 шт.) когенерационных установок, газовой котельной и центральных тепловых пунктов с множеством различных датчиков и управляющих устройств.
- беспроводные датчики контроля температурного режима помещений (более 300 шт.).

Программный комплекс ПолиТЭР реализован на языках программирования C++, R и использует Oracle в качестве базовой СУБД. Поддерживаются открытые и проприетарные протоколы обмена данными с оборудованием различных производителей, конфигурируемая среда визуализации данных, оповещения оператора системы по SMS и электронной почте.

В 2018 г. сенсорная подсистема была существенно расширена IoT-устройствами (установлено более 300 беспроводных датчиков температуры), позволяющими собирать дополнительную информацию о температурном режиме помещений, что позволило существенно оптимизировать систему отопления зданий. Внедрение интеллектуальной системы управления теплоснабжением зданий в кампусе ЮУрГУ позволило в 2018 г. сэкономить около 15 % тепловой энергии.

Основной проблемой эксплуатации измерительного оборудования является периодическое появление ошибочных показаний или данных с наличием возмущений неизвестного характера. Одной из причин является выход оборудования из строя: ввиду устаревания или нарушения условий эксплуатации оборудования ежегодно 10–20 различных устройств подлежат замене. Другой причиной является засорение областей контакта датчика с измерительной средой. Также периодически возникают обрывы линий связи с контрольно-измерительными устройствами, при этом большинство приборов, в особенности, контроллеры, не поддерживают аппаратные архивы данных. Типичной причиной также является некорректная установка оборудования, приводящая к возникновению дополнительных возмущений, связанных с нетиповыми режимами эксплуатации IoT-датчиков в области их установки.

Некорректные данные приводят к следующим основным проблемам:

- ошибочные вычисления в ходе выполнения оптимизационных алгоритмов, построению некорректных характеристик объектов управления и, как следствие, к некорректному принятию стратегических решений;
- некорректное принятие организационно-управляющих решений персоналом при наличии недостоверных отклонений параметров эксплуатации от их номинальных значений;
- некорректное автоматическое управление посредством контроллеров может привести не только к перерасходу энергии, но и к выходу из строя инженерных коммуникаций (например, замораживание системы отопления);
- некорректный расчет потребляемых энергоресурсов или наложение энергоснабжающими организациями на потребителя штрафных санкций за несвоевременное обнаружение и устранение неисправностей.

Своевременное обнаружение, устранение неисправностей оборудования и восстановление данных дает существенный организационно-экономический эффект за счет уменьшения или исключения упущенной выгоды при возникновении указанных проблем.

В описываемом исследовании предложен Модуль очистки данных, расширяющий систему ПолиТЭР и решающий следующие основные задачи:

- оперативное обнаружение аномального поведения датчиков температуры и уведомление оператора о найденных аномалиях;
- оперативное обнаружение пропусков и выбросов в измеряемых данных датчиков и замена таких значений на синтетические правдоподобные данные.

Разработанный Модуль очистки данных располагается в системе ПолиТЭР между уровнем считывания данных и уровнем их использования для аналитических расчетов и представления. При этом отображаемые данные отмечаются как исходные или восстановленные, а доступ к некорректным данным сохраняется для возможности их более глубокого анализа.

2. Обзор работ

Обнаружению выбросов посвящено множество различных методик, многие из которых используются в теории обнаружения отказов оборудования [25, 28]. К базовым подходам относятся методы, определяющие выход за границы допустимого диапазона отклонения фактических значений от значений, рассчитанных по модели объекта [24, 30].

В работе [26] на примере систем мониторинга вентиляции и кондиционирования для определения некорректных показаний используется сочетание трех техник обработки данных: уменьшение размерности коррелирующих данных, сигнализирующих о возникновении выбросов (метод ReliefF и адаптивные генетические алгоритмы); применение расширенного фильтра Калмана для фильтрации шумов и разложения данных во временные ряды; разделение областей корректных и некорректных значений статических и динамических параметров модели объекта с использованием рекурсивного одноклассового метода опорных векторов.

В работе [2] рассматривается задача прогнозного управления теплоснабжением потребителей с помощью нейросети, обученной на статистических данных, получаемых с подсистемы мониторинга. Искусственная нейронная сеть рассматривается в качестве основного инструмента, минимизирующего ошибки, связанные с ручным управлением температурой на выходе из котельной.

Исследование подходов к сбору и анализу данных на предприятиях водоснабжения и водоотведения рассмотрены в работе [3].

Для определения отклонений в работе систем централизованного теплоснабжения в работе [13] предлагается применять метод ближайшего соседа. В данной работе в качестве эталона берутся параметры работы группы схожих автоматизированных индивидуальных тепловых пунктов (АИТП). Неисправностями считаются отклонения показателей работы одного АИТП от показателей группы схожих с ним АИТП, подключенных к одной системе централизованного теплоснабжения, на величину заданного порога.

В работе [15] представлено применение метода главных компонент для идентификации выбросов в данных работы систем отопления, вентиляции и кондиционирования. В работе [12] описан метод восстановления данных, применяемый совместно с методом главных компонент.

Также стоит отметить метод построения балансовых моделей, применяемый к связанным теплоэнергетическим комплексам. Расхождение энергетического или гидравлического баланса в таких моделях свидетельствует о наличии неисправностей в системе или о наличии выбросов [17].

Восстановление данных, утраченных при возникновении выбросов, выполняется путем их расчета на базе методов прогнозирования. Множество работ посвящено прогнозированию параметров систем теплоснабжения с использованием статических и динамических факторных моделей и моделей временных рядов с идентификацией методами авторегрессионного анализа [29], метода опорных векторов [16] и рекуррентных нейронных сетей [5].

Также интерес представляет подход к восстановлению статических и динамических характеристик моделей, искажение которых возникает в результате периодических возмущений. Учет подобных возмущений выполняется путем описания их времени действия индикаторными функциями [8].

При этом необходимо отметить, что особенностью эксплуатационных данных систем теплоснабжения является периодический характер, связанный с цикличностью изменения погодных условий и режимов эксплуатации помещений в зависимости от времени суток. Выбросы в эксплуатационных данных, возникающие при работе различных подсистем, связанных с теплоснабжением, характеризуются резким изменением измеряемого сигнала, длящимся в пределах ограниченного периода времени.

Ввиду этого в данной статье предлагается использовать рекуррентные сети долгой краткосрочной памяти как для определения последовательностей подобных выбросов, так и для восстановления утраченных последовательностей эксплуатационных измерительных данных, носящих циклический характер изменения.

3. Разработка Модуля очистки данных

3.1. Общая архитектура

Общая архитектура Модуля очистки данных представлена на рис. 2. Модуль тиражируется для каждого температурного датчика системы ПолиТЭР и состоит из следующих основных подсистем: Препроцессор, Предиктор, Реконструктор и Детектор аномалий. Препроцессор подготавливает накопленные показания датчиков для дальнейшей обработки. Предиктор обеспечивает искусственную нейронную сеть, которая выполняет прогноз следующего значения датчика на основе его исторических данных. Реконструктор определяет, является ли текущее показание датчика выбросом, и в этом случае заменяет его на синтетическое значение, полученное Предиктором. Детектор аномалий обнаруживает аномальные подпоследовательности в показаниях датчика для последующего уведомления оператора системы ПолиТЭР.

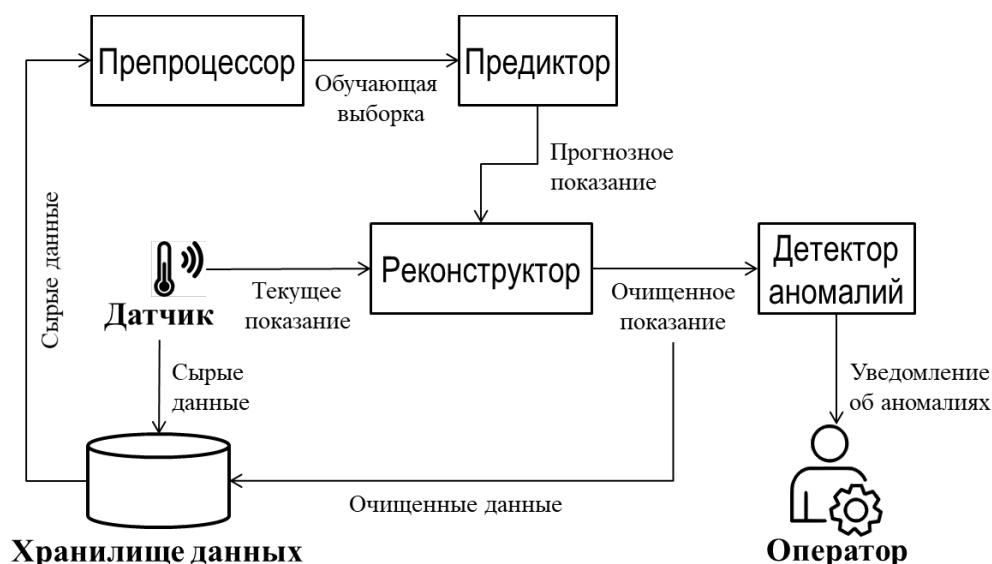


Рис. 2. Архитектура Модуля очистки данных

Технологический цикл Модуля очистки данных для некоторого датчика выглядит следующим образом. Препроцессор выполняет свои действия регулярно с частотой, определяемой оператором системы (типичным случаем является запуск Препроцессора один раз в месяц). Препроцессор извлекает из хранилища данных неочищенные показания датчика, накопленные к текущему моменту, и подготавливает обучающую выборку для нейронной сети подсистемы Предиктор.

Предиктор также запускается регулярно, но в соответствии с частотой снятия показаний датчика (например, один раз в 10 минут), и выполняет следующие действия. Нейронная сеть прогнозирует текущее значение датчика. Если при этом датчик возвратил пустое значение, то оно заменяется на прогнозное. Иначе подсистема Реконструктор выполняет бинарную классификацию значения, полученного от датчика: выброс или норма.

Если текущее значение распознается как выброс, Реконструктор заменяет текущее значение на прогнозное. Модуль очистки данных передает прогнозное значение системе для сохранения в хранилище данных.

В завершении цикла Детектор аномалий определяет, завершает ли текущее значение датчика некоторую аномальную последовательность значений датчика, и уведомляет оператора системы в этом случае.

Далее перечисленные выше подсистемы Модуля очистки данных рассмотрены более детально.

3.2. Препроцессор

Препроцессор выполняет подготовку обучающей выборки для нейронной сети подсистемы Предиктор и состоит из следующих основных подсистем (см. рис. 3): Парсер, Восстановитель, Детектор выбросов и Нормализатор. Технологический цикл работы Препроцессора выглядит следующим образом.

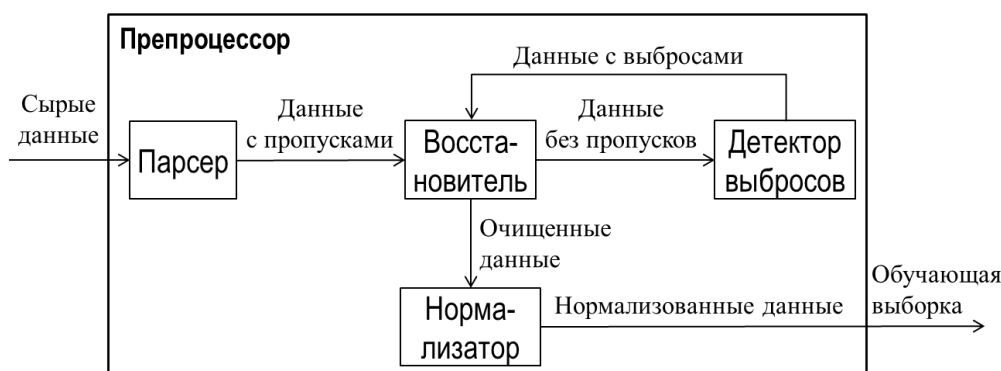


Рис. 3. Архитектура Препроцессора

Сначала Парсер извлекает данные датчика из хранилища данных и преобразует их в формат, пригодный для последующей обработки. Затем Восстановитель заменяет пропущенные значения на правдоподобные синтетические. После этого Детектор выбросов находит в данных точки-выбросы и заменяет их пустыми значениями NULL. Полученные данные повторно подаются на вход Восстановителя, который, таким образом, заменяет выбросы на правдоподобные синтетические значения. На последнем шаге Нормализатор формирует из полученных данных набор нормализованных подпоследовательностей, которые являются обучающей выборкой для нейронной сети подсистемы Предиктор.

Восстановитель использует сезонную авторегрессионную интегрированную модель скользящего среднего с экзогенными регрессорами *SARIMAX* (*Seasonal AutoRegressive Integrated Moving Average with exogenous regressors*) [10], которая часто применяется для прогнозирования значений сезонных временных рядов в различных предметных областях. Свойство сезонности временного ряда подразумевает наличие в данном ряде колебаний, возникающих с некоторой периодичностью. Для применения SARIMAX в данной предметной области необходимо подобрать следующие параметры: p — порядок авторегрессии, d — порядок интегрирования, q — порядок скользящего среднего, P — сезонный порядок авторегрессии, D — порядок сезонного интегрирования, Q — сезонный порядок скользящего среднего, s — период сезонности временного ряда. Применение SARIMAX возможно, если целевой временной ряд является стационарным. Под стационарными вре-

менными рядами понимают такие временные ряды, элементы которых являются случайными величинами с постоянным математическим ожиданием и постоянной дисперсией. Стационарность временного ряда проверяется с помощью следующих двух тестов, применяемых последовательно: расширенный тест Дики—Фуллера (ADF) [11] и тест Квятковского—Филлипса—Шмидта—Шина (KPSS) [21]. Если указанные тесты не подтвердят стационарность временного ряда, то для применения модели SARIMAX необходимо преобразовать элементы ряда в зависимости от его характеристик, например, логарифмировать элементы в случае наличия в ряде мультипликативной сезонности (существенного отличия дисперсии элементов в различных участках ряда).

В более формальном изложении работа Препроцессора выглядит следующим образом. Парсер извлекает из хранилища данных временной ряд T , который представляет собой хронологически упорядоченную последовательность числовых либо пустых значений: $T = (t_1, \dots, t_m)$, где $t_i \in \mathbb{R}$ или $t_i = \text{NULL}$ и число m называется длиной временного ряда. Затем тандем Восстановителя и Детектора выбросов преобразует T таким образом, чтобы $\forall i t_i \in \mathbb{R}$.

Далее Нормализатор подготавливает обучающую выборку, каждый элемент которой представляет собой пару «последовательность показаний датчика» и «прогнозное значение датчика». В качестве первого элемента указанной пары берется нормализованная подпоследовательность показаний датчика фиксированной длины, в качестве второго элемента — одно нормализованное показание датчика, следующее за данной подпоследовательностью.

Подпоследовательностью $T_{i,n}$ временного ряда T назовем непрерывный промежуток значений указанного ряда, состоящий из n элементов и начинающийся с позиции i : $T_{i,n} = (t_i, \dots, t_{i+n-1})$, $1 \leq i \leq m - n + 1$. Используя минимаксную нормализацию, для подпоследовательности $T_{i,n}$ ее нормализованная версия $\tilde{T}_{i,n}$ вычисляется как $\tilde{T}_{i,n} = (\tilde{t}_i, \dots, \tilde{t}_{i+n-1})$, где $\tilde{t}_i = \frac{t_i - t_{\min}}{t_{\max} - t_{\min}}$. Для нормализованной подпоследовательности $\tilde{T}_{i,n}$ элемент \tilde{t}_{n+1} рассматривается как ее прогнозное значение. Для дальнейшего использования введем обозначение S_T^n набора нормализованных подпоследовательностей временного ряда T , имеющих длину n ($n \ll m$), и за P обозначим набор соответствующих этим подпоследовательностям прогнозов.

Длина подпоследовательности n ($n \ll m$) является параметром Модуля очистки и вычисляется как $n = \text{frequency} \cdot \text{horizon}$, где *frequency* — частота датчика, а *horizon* — исторический горизонт (длина временного интервала в прошлом), используемый Предиктором и подбираемый оператором SCADA-системы «ПолиТЭР». Например, типичная частота снятия показаний температурного датчика 4 раза в час, исторический горизонт выбран равным 12 часам, тогда длина подпоследовательности, используемая Модулем очистки, составляет $n=48$.

3.3. Предиктор

Предиктор представляет собой рекуррентную нейронную сеть (Recurrent Neural Network, RNN) со слоем долгой краткосрочной памяти (Long Short-Term Memory, LSTM) [14]. Обучение рекуррентной нейронной сети выполняется на наборе данных, подготовленном Препроцессором. При использовании нейронная сеть принимает в качестве входных данных подпоследовательность реальных значений датчика, предшествующих текущему значению, и выдает прогнозное значение.

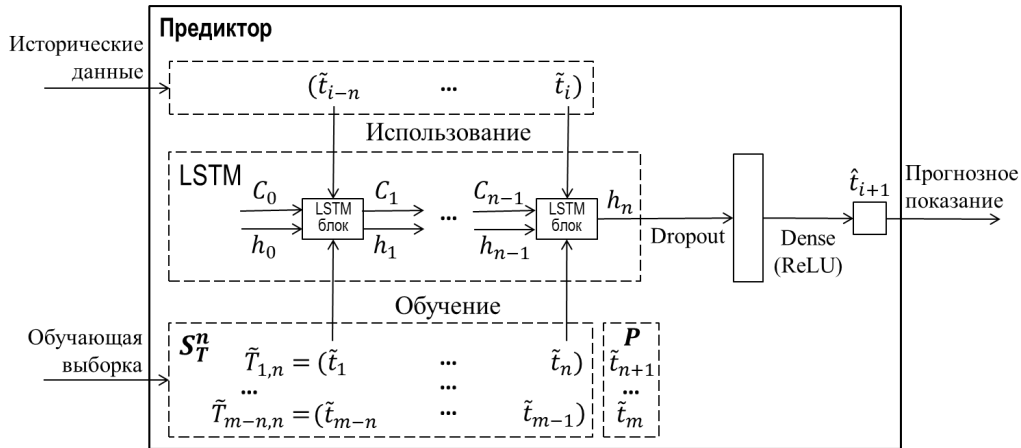


Рис. 4. Архитектура Предиктора

Архитектура Предиктора представлена на рис. 4. Рекуррентная нейронная сеть состоит из следующих слоев. Слой LSTM состоит из однотипных LSTM-блоков, где количество блоков равно длине подпоследовательности n , выбранной на этапе предварительной обработки. Работая вместе, блоки LSTM создают вектор-столбец h , т.н. скрытое состояние. Длина этого вектора — параметр, определяемый оператором SCADA-системы «ПолиТЭР». Слой Dropout случайным образом деактивирует predetermined долю нейронов в векторе h для предотвращения переобучения нейронной сети. Доля деактивируемых нейронов также является параметром, который определяется оператором системы, с типичным значением 20 %. Слой Dense применяет выпрямленный линейный блок (Rectified Linear Unit, ReLU) в качестве функции активации для преобразования данных в одно прогнозное значение \hat{t}_{i+1} .

Каждый LSTM-блок состоит из состояния блока и нескольких слоев фильтра. Состояние блока — это вектор, который несет информацию из предыдущих моментов и будет проходить через всю цепочку LSTM-блоков. LSTM-блок имеет три слоя фильтра («вентили»), которые используются для контроля потоков информации на входах и на выходах памяти данных блоков: входной фильтр, фильтр забывания и выходной фильтр, которые регулируют объем данных, которые должны быть сохранены, забыты и доставлены на выход соответственно.

3.4. Реконструктор

Реконструктор получает на входе непустое текущее показание датчика и проверяет, существенно ли оно отличается от денормализованного синтетического значения, полученного с помощью Предиктора. Если это так, то реальное значение заменяется на синтетическое, которое в итоге сохраняется в хранилище данных.

Реализация Реконструктора основана на использовании распределения вероятностей ошибки прогнозирования [22]. В соответствии с этим подходом определяется порог различия $\varepsilon > 0$ между реальным показанием датчика t_{i+1} и прогнозным синтетическим значением датчика \hat{t}_{i+1} . Если $|t_{i+1} - \hat{t}_{i+1}| \geq \varepsilon$, то t_{i+1} представляет собой выброс. Порог различия вычисляется как $\varepsilon = \mu + k\sigma$, где μ — среднее значение ошибок прогнозирования, σ — стандартное отклонение ошибок прогнозирования, а $k > 0$ — параметр, определяемый оператором системы ПолиТЭР (с типичным значением $k=3$). Для некоторой подпоследовательности из тестовой выборки подсистемы Предиктор ошибка прогнозирования

вычисляется как модуль разности между последней точкой данной подпоследовательности и соответствующим прогнозным значением, которое было выдано подсистемой Предиктор.

3.5. Детектор аномалий

Детектор аномалий проверяет, является ли аномалией некая подпоследовательность показаний датчика, которая заканчивается его текущим значением, и в случае положительного ответа формирует уведомление для оператора системы ПолиТЭР. В качестве длины указанной подпоследовательности рассматривается набор значений, которые соответствуют значимым временным интервалам в сфере управления отоплением и вычисляются в зависимости от частоты снятия показаний датчика. Например, если частота датчика составляет 4 раза в час, и оператору системы ПолиТЭР необходимо уведомление об аномалиях за прошедшие 12 час., 24 час. и 48 час., то в качестве параметра Детектора аномалий оператор должен задать следующие длины подпоследовательностей: 48, 96 и 182 соответственно.

В реализации Детектора аномалий нами используется концепция диссонанса [18, 27]. Диссонанс, в отличие от большинства алгоритмов обнаружения аномалий, требует только один интуитивно понятный параметр: длина аномальной подпоследовательности [19] и формально определяется следующим образом. Две подпоследовательности $T_{i,n}$ и $T_{j,n}$ ряда T не являются тривиальными совпадениями друг друга, если $\exists T_{p,n} \in S_T^n, i < p < j: ED(T_{i,n}, T_{j,n}) < ED(T_{i,n}, T_{p,n})$, где $ED(\cdot, \cdot)$ обозначает евклидово расстояние. Пусть M_C обозначает множество подпоследовательностей, которые не являются тривиальным совпадением подпоследовательности $C \in S_T^n$. Тогда подпоследовательность $D \in S_T^n$ назовем наиболее значимым диссонансом в T , если $\forall C \in S_T^n \min(ED(D, M_D)) > \min(ED(C, M_C))$.

Иными словами, подпоследовательность ряда является диссонансом, если евклидово расстояние от нее до ближайшей подпоследовательности, не являющейся ее тривиальным совпадением, является наибольшим. Степень значимости диссонансов определяется следующим образом. Подпоследовательность $D \in S_T^n$ называется наиболее значимым k -м диссонансом в T , если евклидово расстояние от нее до k -й ближайшей подпоследовательности, не являющейся ее тривиальным совпадением, является наибольшим.

Таким образом, Детектор аномалий формирует уведомление для оператора системы ПолиТЭР, если подпоследовательность предопределенной длины, заканчивающаяся текущим показанием датчика, является диссонансом со степенью значимости не ниже k , где k является параметром, также определяемым оператором системы.

3.6. Инструменты и технологии реализации

Препроцессор реализован с помощью библиотек языка Python следующим образом. Парсер использует стандартные библиотеки *openpyxl* и *pandas*. Детектор выбросов основан на алгоритмах библиотеки *adtk (Anomaly Detection Toolkit)* [7]. Модель SARIMAX, используемая Восстановителем, реализована в стандартной библиотеке *statsmodels*. Нормализатор реализован с помощью стандартной библиотеки *sklearn*.

Предиктор реализован на основе библиотеки *Keras* [20] и фреймворка *TensorFlow* [4]. Для обучения рекуррентной нейронной сети использовались подпоследовательности

длины $n=48$, соответствующие 12 час. работы температурного датчика. Скрытое состояние рекуррентной нейронной сети было взято как вектор-столбец длины $|h|=32$. Для обучения сети использовались следующие общепринятые параметры: функция потерь MSE (Mean Square Error, среднеквадратическая ошибка), оптимизатор обучения Adam, количество эпох и размер пакета — 15 и 32 соответственно.

Для реализации Детектора аномалий использована библиотека *MatrixProfile* для языка программирования Python [9].

4. Вычислительные эксперименты

4.1. Набор данных

Для оценки предложенного подхода нами были проведены вычислительные эксперименты на реальных данных, взятых из хранилища данных SCADA-системы «ПолиТЭР». В качестве данных использовались показания температурного датчика, установленного в аудитории для потоковых лекций, за 2018 г. (частота снятия показаний датчика составляет 4 раза в час).

После обработки указанных данных Препроцессором первые 42 недели года (80 %) показаний были использованы в двух качествах: как выборка для подбора параметров модели SARIMAX Восстановителя и как обучающая выборка рекуррентной нейронной сети подсистемы Предиктор. Оставшиеся 8 недель года (20 %) показаний датчика моделировали работу модуля очистки данных в штатном режиме и использовались в качестве тестовой выборки.

4.2. Подбор параметров модели SARIMAX

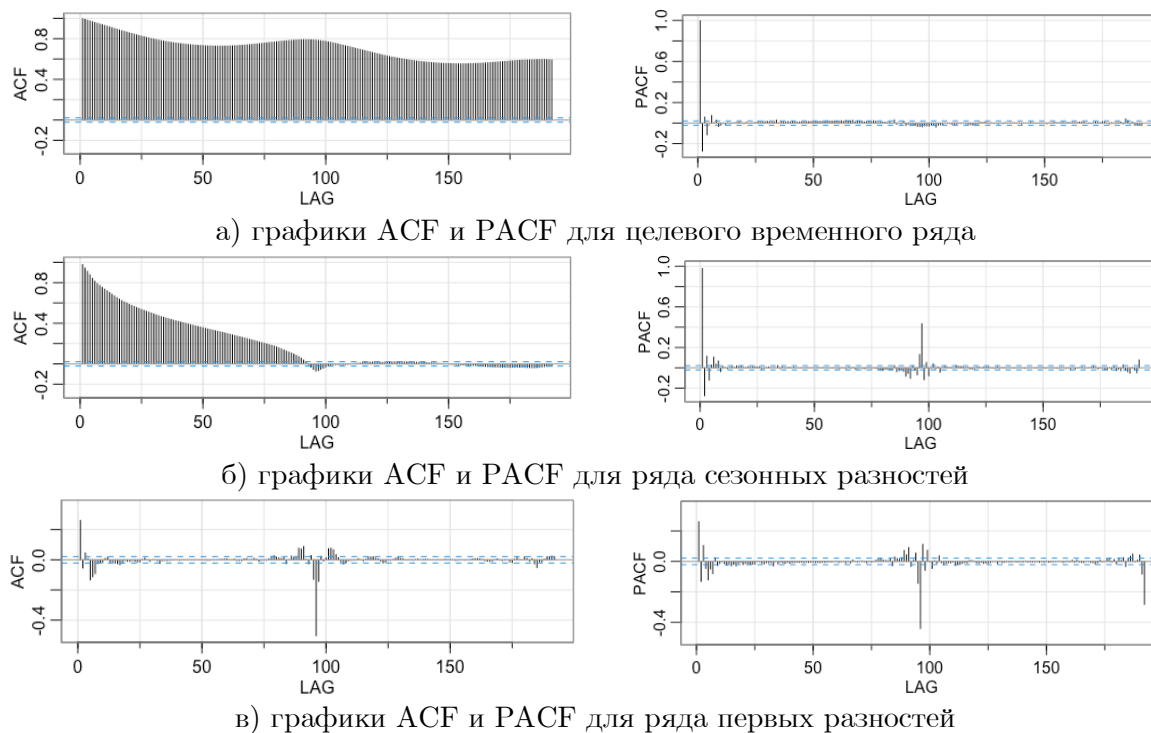


Рис. 5. Коррелограммы исходного набора данных

Проверка взятого для исследования набора данных с помощью тестов ADF и KPSS подтвердила его нестационарность. Для приведения временного ряда к стационарному виду нами были последовательно применены техники вычитания первых разностей для удаления тренда и вычитания сезонных разностей для удаления сезонности [10]. Соответствующие графики функций автокорреляции (ACF) и частичной автокорреляции (PACF) приведены на рис. 5. Итоговые графики на рис. 5в позволяют подобрать следующие диапазоны для параметров модели SARIMAX: порядок авторегрессии $p \in [0, 7]$, порядок интегрирования $d = 1$, порядок скользящего среднего $q \in [0, 7]$, сезонный порядок авторегрессии $P = 1$, порядок сезонного интегрирования $D = 1$, сезонный порядок скользящего среднего $Q = 0$, период сезонности временного ряда $s = 96$.

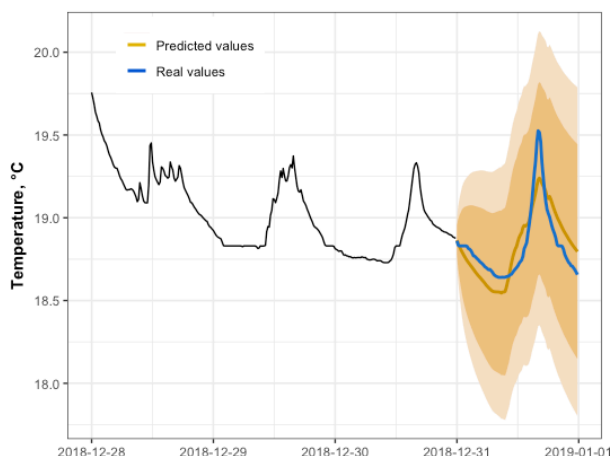
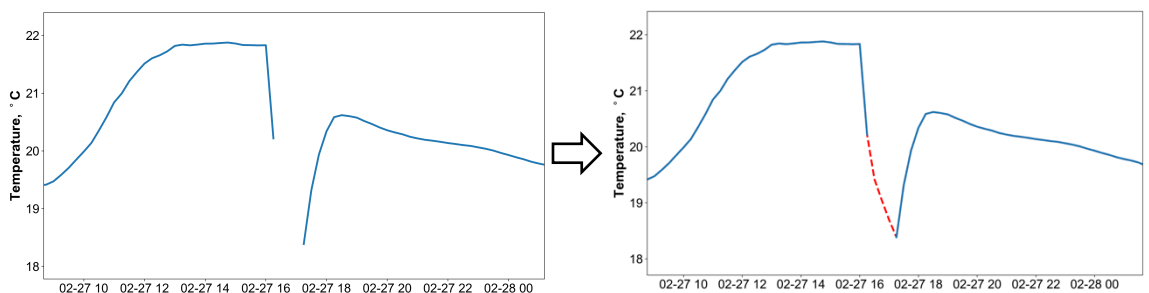
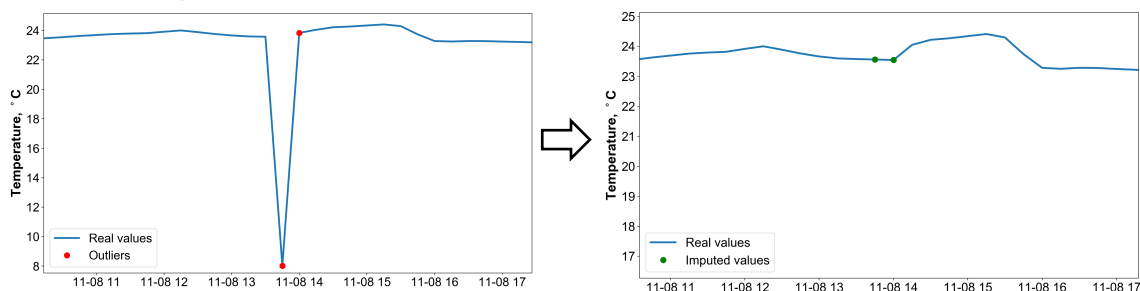


Рис. 6. Прогноз на один день, полученный с помощью модели SARIMAX(6,1,5)(0,1,1)₉₆



а) Восстановление пропущенных значений временного ряда



б) Замена выбросов временного ряда на правдоподобные синтетические значения

Рис. 7. Результаты работы Восстановителя на основе модели SARIMAX(6,1,5)(0,1,1)₉₆

Среди всех полученных моделей для штатной работы Модуля очистки данных была выбрана модель с параметрами SARIMAX($p = 6, d = 1, q = 5$)($P = 0, D = 1, Q = 1$) $s = 96$, поскольку имеет наименьшие значения информационных критериев Акаике (AIC) [6] и Шварца (BIC) [23] по сравнению с другими моделями и ряд остатков (разность между

фактическими и прогнозными значениями) обладает свойствами несмещенности, стационарности и неавтокоррелированности. На рис. 6 представлен прогноз на один день, а также построены 80 % и 95 % доверительные интервалы прогноза. Ошибка прогноза составила по мере $RMSE = 0.119$.

Поскольку подобранная нами модель SARIMAX довольно точно предсказывает значения температурного датчика, следовательно, она может быть использована для замены пропущенных значений на правдоподобные синтетические значения. Примеры работы Восстановителя представлены на рис. 7.

4.3. Оценка точности восстановления данных

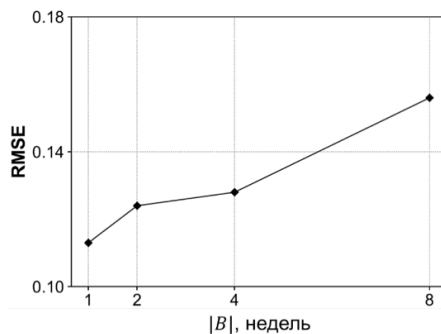
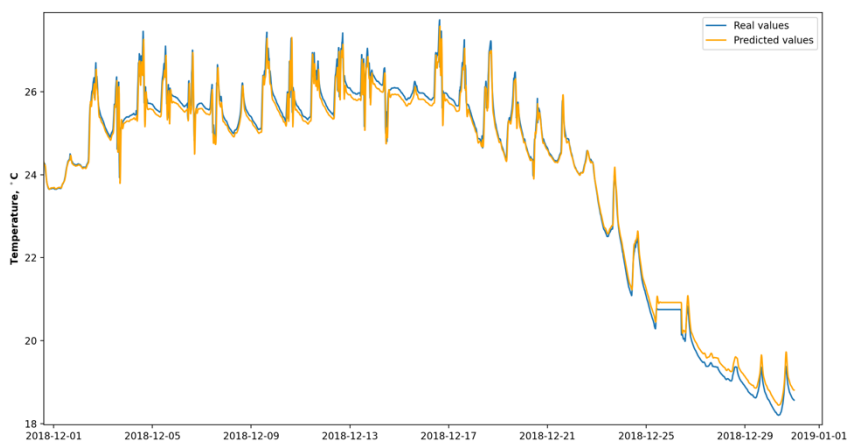
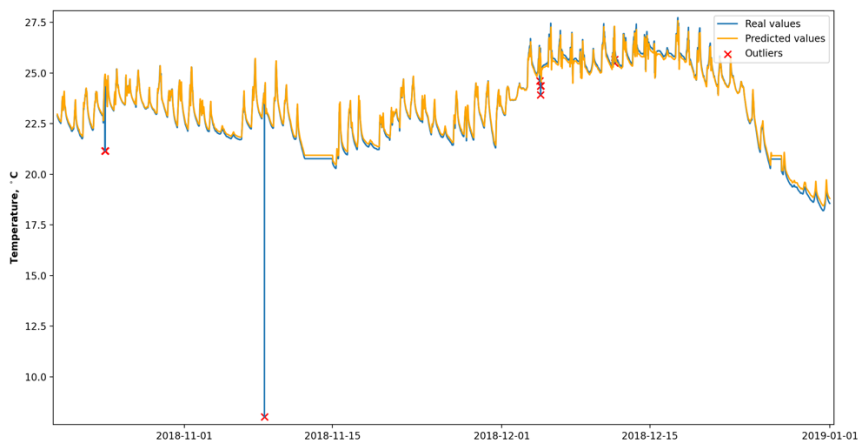


Рис. 8. Точность работы Модуля очистки данных



а) $|B| = 4$ недели



б) $|B| = 8$ недель

Рис. 9. Моделирование работы Модуля очистки данных

В экспериментах показания указанного датчика, синтезированные модулем очистки данных, сравнивались с реальными, полученными из хранилища данных, и оценивалась точность соответствия. В качестве меры точности использовалась среднеквадратичная ошибка (RMSE, Root Mean Square Error), определяемая следующим образом: $RMSE = \sqrt{\frac{1}{|B|} \sum_{i=1}^{|B|} (t_i - \hat{t}_i)^2}$, где t_i и \hat{t}_i — реальное и синтетическое значения датчика соответственно, а $|B|$ обозначает длину блока показаний датчика, на которых выполнялось сравнение. Результаты экспериментов, в которых длина блока варьировалась от одной недели до двух месяцев, представлены на рис. 8. Как видно, разработанный модуль обеспечивает относительно высокую и стабильную точность.

На рис. 9 показаны примеры моделирования работы модуля для различных длин блока показаний датчика. Можно видеть, что модуль адекватно предсказывает нормальные значения и обнаруживает точечные выбросы в данных.

4.4. Обнаружение аномалий в данных

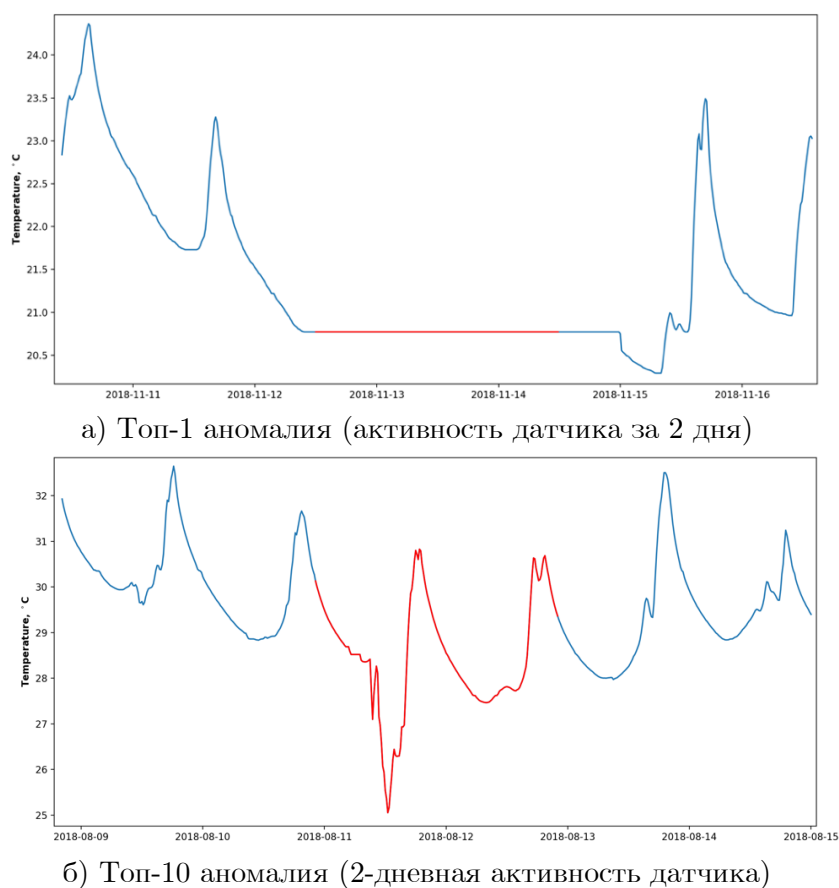


Рис. 10. Аномалии, обнаруженные при моделировании работы Модуля очистки данных

На рис. 10 показан пример двух аномалий, обнаруженных при моделировании работы модуля очистки данных и соответствующих двухдневной активности вышеуказанного датчика. Первая аномалия представляет собой топ 1 диссонанс, найденный в тестовом наборе данных, которая может указывать на временный сбой в работе датчика. Вторая аномалия представляет собой топ 10 диссонанс и может указывать на быстрое снижение

температуры в лекционном зале из-за интенсивной вентиляции ввиду большого количества открытых окон в жаркий день. В любом случае, обнаруженные аномалии являются предметом реакции оператора.

Заключение

В статье затронута проблема очистки данных, поступающих с температурных датчиков в интеллектуальных системах отопления умных домов. Рассмотрен случай SCADA-системы ПолиТЭР, установленной в Южно-Уральском государственном университете (Челябинск) для управления отоплением университетского кампуса. Представлены архитектура и принципы реализации модуля очистки данных, разработанного и внедренного авторами в систему ПолиТЭР. Модуль очистки данных тиражируется для каждого температурного датчика системы ПолиТЭР и состоит из следующих подсистем: Препроцессор, Предиктор, Реконструктор и Детектор аномалий.

Препроцессор подготавливает обучающий набор данных для рекуррентной нейронной сети подсистемы Предиктор и реализуется с помощью следующих подсистем. Парсер извлекает данные датчика из хранилища данных и преобразует их для последующей обработки. Затем Восстановитель заменяет пропущенные значения на правдоподобные синтетические. После этого Детектор выбросов находит в данных точки-выбросы и заменяет их пустыми значениями, и полученные данные повторно обрабатываются Восстановителем. На последнем шаге Нормализатор формирует из полученных данных набор нормализованных подпоследовательностей, которые являются обучающей выборкой для рекуррентной нейронной сети подсистемы Предиктор.

Предиктор обеспечивает рекуррентную нейронную сеть долгой краткосрочной памяти, которая обучается на данных, подготовленных Препроцессором, получает на входе подпоследовательность реальных значений датчика, оканчивающихся его текущим значением, и выдает прогнозное значение датчика. Реконструктор получает на входе текущее значение датчика и проверяет, существенно ли оно отличается от синтетического значения, полученного с помощью Предиктора. Если это так, то реальное значение заменяется на синтетическое, которое в итоге сохраняется в хранилище данных. Детектор аномалий проверяет, является ли аномалией подпоследовательность показаний датчика, оканчивающаяся его текущим значением, и в случае положительного ответа формирует уведомление для оператора системы ПолиТЭР.

Проведены вычислительные эксперименты на реальных данных датчиков системы ПолиТЭР, показавшие высокую точность работы Модуля очистки данных.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 20-07-00140) и Министерства образования и науки РФ (государственное задание FENU-2020-0022).

Литература

1. Басалаев А.А. Автоматизированный энергоменеджмент теплоэнергетического комплекса университетского городка // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2015. Т. 15, № 4. С. 26–32. DOI: 10.14529/ctcr150403.

2. Нетбай Г.В., Онискив В.Д., Столбов В.Ю., Каримов Р.Р. Прогнозное управление локальной городской системой теплоснабжения на основе нейросетевого моделирования // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2019. № 3(47). С. 28–38. DOI: 10.14529/ctcr200303.
3. Ясир Ш., Кравец А.Г., Анохин А.О., Пивоваров В.В., Астанков А.А. Сбор и анализ гетерогенных данных в управлении услугами ЖКХ по водоснабжению и водоотведению // Прикаспийский журнал: управление и высокие технологии. 2015. Т. 15, № 4. С. 26–32. DOI: 10.14529/ctcr150403.
4. Abadi M., Barham P., Chen J., et al. TensorFlow: A System for Large-Scale Machine Learning // Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, OSDI'16 (Berkeley, CA, United States, November 2016). 2016. P. 265–283. URL: <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>.
5. Ahmad T., Chen H., Huanga Y. Short-Term Energy Prediction for District-Level Load Management Using Machine Learning Based Approaches // Energy Procedia. 2019. Vol. 158. P. 3331–3338. DOI: 10.1016/j.egypro.2019.01.967.
6. Akaike H. A new look at the statistical model identification // IEEE Transactions on Automatic Control. 1974. Vol. 19, no. 6. P. 716–723. DOI: 10.1109/TAC.1974.1100705.
7. Anomaly Detection Toolkit, User Guide. URL: <https://arundo-adtk.readthedocs-hosted.com/en/stable/userguide.html> (дата обращения: 02.09.2020).
8. Basalaev A., Tochilkin M., Shnayder D. Enhancing room thermal comfort conditions modeling in buildings through schedule-based indicator functions for possible variable thermal perturbation inputs // Proceedings of 2019 International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM 2019 (Chelyabinsk, Russia, March 25–29, 2019). 2019. P. 1–8. DOI: 10.1109/ICIEAM.2019.8742907.
9. Benschoten A.H.V., Ouyang A., Bischoff F., Marrs T. MPA: a novel cross-language API for time series analysis // Journal of Open Source Software. 2020. Vol. 5, no. 49. DOI: 10.21105/joss.02179.
10. Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. Time Series Analysis: Forecasting and Control. Wiley, 2015. 712 p.
11. Dickey D.A., Fuller W.A. Distribution of the estimators for autoregressive time series with a unit root // Journal of the American Statistical Association. 1979. Vol. 74. P. 427–431.
12. Dunia R., Joe Qin S. Joint diagnosis of process and sensor faults using principal component analysis // Control Engineering Practice. 1998. Vol. 6, no. 4. P. 457–469. DOI: 10.1016/S0967-0661(98)00027-6.
13. Farouq Sh., Byttner S., Bouguelia M.-R., et al. Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach // Engineering Applications of Artificial Intelligence. 2020. Vol. 90. P. 1–16. DOI: 10.1016/j.engappai.2020.103492.
14. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997. Vol. 9, no. 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
15. Hu Y., Chen H., Li G., et al. A statistical training data cleaning strategy for the PCA-based chiller sensor fault detection, diagnosis and data reconstruction method // Energy and Buildings. 2016. Vol. 112. P. 270–278. DOI: 10.1016/j.enbuild.2015.11.066.
16. Idowu S., Saguna S., Åhlund C., Schelén O. Applied machine learning: Forecasting heat load in district heating system // Energy and Buildings. 2016. Vol. 133. P. 478–488. DOI: 10.1016/j.enbuild.2016.09.068.

17. Jha K. Minimal loop extraction for leak detection in water pipe network // Proceedings of 2012 1st International Conference on Recent Advances in Information Technology, RAIT 2012 (Dhanbad, India, March 15–17, 2012). IEEE Computer Society, 2012. P. 687–693. DOI: 10.1109/RAIT.2012.6194578.
18. Keogh E.J., Lin J., Fu A.W. HOT SAX: efficiently finding the most unusual time series subsequence // Proceedings of the 5th IEEE International Conference on Data Mining, ICDM 2005 (Houston, Texas, USA, November 27–30, 2005). IEEE Computer Society, 2005. P. 226–233. DOI: 10.1109/ICDM.2005.79.
19. Keogh E.J., Lonardi S., Ratanamahatana C.A. Towards parameter-free data mining // Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Seattle, Washington, USA, August 22–25, 2004). ACM, 2004. P. 206–215. DOI: 10.1145/1014052.1014077.
20. Keras Developer Guides. URL: <https://keras.io/guides/> (дата обращения: 02.09.2020).
21. Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? // Journal of Econometrics. 1992. Vol. 54, no. 1. P. 154–178. DOI: 10.1016/0304-4076(92)90104-Y.
22. Malhotra P., Vig L., Shroff G.M., Agarwal P. Long Short Term Memory Networks for anomaly detection in time series // Proceedings of the 23rd European Symposium on Artificial Neural Networks, ESANN 2015 (Bruges, Belgium, April 22–24, 2015). 2015. P. 89–94. URL: <http://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2015-56.pdf>.
23. Schwarz G. Estimating the dimension of a model // The Annals of Statistics. 1978. Vol. 6, no. 2. P. 461–464. DOI: 10.1214/aos/1176344136.
24. Turner W.J.N., Staino A., Basu B. Residential HVAC fault detection using a system identification approach // Energy and Buildings. 2017. Vol. 151. P. 1–17. DOI: 10.1016/j.enbuild.2017.06.008.
25. Venkatasubramanian V. Process Fault Detection and Diagnosis: Past, Present and Future // IFAC Proceedings Volumes. 2001. Vol. 34, no. 27. P. 1–13. DOI: 10.1016/S1474-6670(17)33563-2.
26. Yan K., Ji Zh., Shen W. Online fault detection methods for chillers combining extended Kalman filter and Recursive One-class SVM // Neurocomputing. 2017. Vol. 228. P. 205–212. DOI: 10.1016/j.neucom.2016.09.076.
27. Yankov D., Keogh E.J., Rebbapragada U. Disk aware discord discovery: finding unusual time series in terabyte sized datasets // Knowledge and Information Systems. 2008. Vol. 17, no. 2. P. 241–262. DOI: 10.1007/s10115-008-0131-9.
28. Zhao Y., Li T., Zhang X., Zhang C. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future // Renewable and Sustainable Energy Reviews. 2019. Vol. 109. P. 85–101. DOI: 10.1016/j.rser.2019.04.021.
29. Zhao Y., Zhang C., Zhang Y., et al. A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis // Energy and Built Environment. 2020. Vol. 1, no. 2. P. 149–164. DOI: 10.1016/j.enbenv.2019.11.003.
30. Zimmerman N., Dahlquist E., Kyprianidis K. Towards On-line Fault Detection and Diagnostics in District Heating Systems // Energy Procedia. 2017. Vol. 105. P. 1960–1966. DOI: 10.1016/j.egypro.2017.03.567.
31. Zymbler M., Kraeva Ya., Latypova E., Kumar S., Shnayder D., Basalaev A. Cleaning Sensor Data in Smart Heating Control System // Proceedings — 2020 Global Smart Industry Conference, GloSIC 2020 (Chelyabinsk, Russia, November 17–19, 2020). IEEE, 2020. P. 375–381. DOI: 10.1109/GloSIC50886.2020.9267813.

Цымблер Михаил Леонидович, д.ф.-м.н., доцент, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Краева Яна Александровна, преподаватель, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Латыпова Елизавета Альбертовна, студент, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Иванова Елена Владимировна, к.ф.-м.н., кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Шнайдер Дмитрий Александрович, д.т.н., доцент, кафедра «Автоматика и управление», Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Басалаев Александр Анатольевич, к.т.н., кафедра «Автоматика и управление», Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

DOI: 10.14529/cmse210302

CLEANING SENSOR DATA IN INTELLIGENT HEATING CONTROL SYSTEM

© 2021 M.L. Zymbler, Ya.A. Kraeva, E.A. Latypova, E.V. Ivanova,
D.A. Shnyder, A.A. Basalaev

South Ural State University (pr. Lenina 76, Chelyabinsk, 454080 Russia)

E-mail: mzym@susu.ru, kraevaya@susu.ru, latypovaea@susu.ru,

elena.ivanova@susu.ru, shneiderda@susu.ru, basalaevaa@susu.ru

Received: 03.09.2020

Sometimes, smart heating control applications are partially equipped with missing values and outliers in the sensor data due to software/hardware failures/human errors. To provide an effective analysis and decision-making, erroneous sensor data should be cleaned by imputation of missing values and smoothing outliers. In this paper, we present a case of the Smart Heating Control System (SHCS) installed in the South Ural State University, and describe the structure and development principles of Data Cleaning Module (DCM) of the system. We implement DCM through data mining and neural network technologies as a set of the following subsystems. The preprocessor extracts raw data from the system's data warehouse and prepares a training data for further processing. Predictor provides Recurrent Neural Network (RNN) to forecast the next value of a sensor based on its historical data. Reconstructor determines if the current value of a sensor is an outlier, and if so, imputes it by the synthetic value from Predictor. Finally, Anomaly Detector subsystem discovers anomalous sequences in the sensor data. In the experiments on the real sensor data, DCM showed relatively high and stable accuracy as well as adequate detection of anomalies.

Keywords: heating systems, control systems, cleaning, process control, temperature sensors, flowmeters, temperature measurement.

FOR CITATION

Zymbler M.L., Kraeva Ya.A., Latypova E.A., Ivanova E.V., Shnyder D.A., Basalaev A.A. Cleaning Sensor Data in Intelligent Heating Control System. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2021. Vol. 10, no. 3. P. 16–36. (in Russian) DOI: 10.14529/cmse210302.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Basalaev A. Automated energy management for heat and power system of university campus. Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics. 2015. Vol. 15, no. 4. P. 26–32. DOI: 10.14529/ctcr150403.
2. Netbay G.V., Oniskiv V.D., Stolbov V.Yu., Karimov R.R. Management of a local urban heat supply system based on neural network modeling taking into account the weather forecast. Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics. 2019. Vol. 20, no. 3. P. 28–38. DOI: 10.14529/ctcr200303.
3. Yasir Sh., Kravets A.G., Anokhin A.O., Pivovarov V.V., Astankov A.A. Collection and analysis of heterogeneous data in the management of housing services on water supply and water disposal. Caspian Journal: Control and High Technologies. 2019. Vol. 3, no. 47. P. 26–32. DOI: 10.14529/ctcr150403.
4. Abadi M., Barham P., Chen J., et al. TensorFlow: A System for Large-Scale Machine Learning. Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, OSDI'16 (Berkeley, CA, United States, November 2016). 2016. P. 265–283. URL: <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>.
5. Ahmad T., Chen H., Huang Y. Short-Term Energy Prediction for District-Level Load Management Using Machine Learning Based Approaches. Energy Procedia. 2019. Vol. 158. P. 3331–3338. DOI: 10.1016/j.egypro.2019.01.967.
6. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974. Vol. 19, no. 6. P. 716–723. DOI: 10.1109/TAC.1974.1100705.
7. Anomaly Detection Toolkit, User Guide. URL: <https://arundo-adtk.readthedocs-hosted.com/en/stable/userguide.html> (accessed: 02.09.2020).
8. Basalaev A., Tochilkin M., Shnyder D. Enhancing room thermal comfort conditions modeling in buildings through schedule-based indicator functions for possible variable thermal perturbation inputs. Proceedings of 2019 International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM 2019 (Chelyabinsk, Russia, March 25–29, 2019). 2019. P. 1–8. DOI: 10.1109/ICIEAM.2019.8742907.
9. Benschoten A.H.V., Ouyang A., Bischoff F., Marrs T. MPA: a novel cross-language API for time series analysis. Journal of Open Source Software. 2020. Vol. 5, no. 49. DOI: 10.21105/joss.02179.
10. Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. Time Series Analysis: Forecasting and Control. Wiley, 2015. 712 p.
11. Dickey D.A., Fuller W.A. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association. 1979. Vol. 74. P. 427–431.
12. Dunia R., Joe Qin S. Joint diagnosis of process and sensor faults using principal component analysis. Control Engineering Practice. 1998. Vol. 6, no. 4. P. 457–469. DOI: 10.1016/S0967-0661(98)00027-6.
13. Farouq Sh., Byttner S., Bouguelia M.-R., et al. Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach. Engineering Applications of Artificial Intelligence. 2020. Vol. 90. P. 1–16. DOI: 10.1016/j.engappai.2020.103492.

14. Hochreiter S., Schmidhuber J. Long Short-Term Memory. *Neural Computation*. 1997. Vol. 9, no. 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
15. Hu Y., Chen H., Li G., et al. A statistical training data cleaning strategy for the PCA-based chiller sensor fault detection, diagnosis and data reconstruction method. *Energy and Buildings*. 2016. Vol. 112. P. 270–278. DOI: 10.1016/j.enbuild.2015.11.066.
16. Idowu S., Saguna S., Åhlund C., Schelén O. Applied machine learning: Forecasting heat load in district heating system. *Energy and Buildings*. 2016. Vol. 133. P. 478–488. DOI: 10.1016/j.enbuild.2016.09.068.
17. Jha K. Minimal loop extraction for leak detection in water pipe network. *Proceedings of 2012 1st International Conference on Recent Advances in Information Technology, RAIT 2012 (Dhanbad, India, March 15–17, 2012)*. IEEE Computer Society, 2012. P. 687–693. DOI: 10.1109/RAIT.2012.6194578.
18. Keogh E.J., Lin J., Fu A.W. HOT SAX: efficiently finding the most unusual time series subsequence. *Proceedings of the 5th IEEE International Conference on Data Mining, ICDM 2005 (Houston, Texas, USA, November 27–30, 2005)*. IEEE Computer Society, 2005. P. 226–233. DOI: 10.1109/ICDM.2005.79.
19. Keogh E.J., Lonardi S., Ratanamahatana C.A. Towards parameter-free data mining. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Seattle, Washington, USA, August 22–25, 2004)*. ACM, 2004. P. 206–215. DOI: 10.1145/1014052.1014077.
20. Keras Developer Guides. URL: <https://keras.io/guides/> (accessed: 02.09.2020).
21. Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*. 1992. Vol. 54, no. 1. P. 154–178. DOI: 10.1016/0304-4076(92)90104-Y.
22. Malhotra P., Vig L., Shroff G.M., Agarwal P. Long Short Term Memory Networks for anomaly detection in time series. *Proceedings of the 23rd European Symposium on Artificial Neural Networks, ESANN 2015 (Bruges, Belgium, April 22–24, 2015)*. 2015. P. 89–94. URL: <http://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2015-56.pdf>.
23. Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978. Vol. 6, no. 2. P. 461–464. DOI: 10.1214/aos/1176344136.
24. Turner W.J.N., Staino A., Basu B. Residential HVAC fault detection using a system identification approach. *Energy and Buildings*. 2017. Vol. 151. P. 1–17. DOI: 10.1016/j.enbuild.2017.06.008.
25. Venkatasubramanian V. Process Fault Detection and Diagnosis: Past, Present and Future. *IFAC Proceedings Volumes*. 2001. Vol. 34, no. 27. P. 1–13. DOI: 10.1016/S1474-6670(17)33563-2.
26. Yan K., Ji Zh., Shen W. Online fault detection methods for chillers combining extended Kalman filter and Recursive One-class SVM. *Neurocomputing*. 2017. Vol. 228. P. 205–212. DOI: 10.1016/j.neucom.2016.09.076.
27. Yankov D., Keogh E.J., Rebbapragada U. Disk aware discord discovery: finding unusual time series in terabyte sized datasets. *Knowledge and Information Systems*. 2008. Vol. 17, no. 2. P. 241–262. DOI: 10.1007/s10115-008-0131-9.
28. Zhao Y., Li T., Zhang X., Zhang C. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renewable and Sustainable Energy Reviews*. 2019. Vol. 109. P. 85–101. DOI: 10.1016/j.rser.2019.04.021.

29. Zhao Y., Zhang C., Zhang Y., et al. A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. *Energy and Built Environment*. 2020. Vol. 1, no. 2. P. 149–164. DOI: 10.1016/j.enbenv.2019.11.003.
30. Zimmerman N., Dahlquist E., Kyprianidis K. Towards On-line Fault Detection and Diagnostics in District Heating Systems. *Energy Procedia*. 2017. Vol. 105. P. 1960–1966. DOI: 10.1016/j.egypro.2017.03.567.
31. Zymbler M., Kraeva Ya., Latypova E., Kumar S., Shnayder D., Basalaev A. Cleaning Sensor Data in Smart Heating Control System. *Proceedings — 2020 Global Smart Industry Conference, GloSIC 2020 (Chelyabinsk, Russia, November 17–19, 2020)*. IEEE, 2020. P. 375–381. DOI: 10.1109/GloSIC50886.2020.9267813.