

В. Н. Захаров, Д. Ю. Ковалев, С. А. Ступников, И. А. Шанин (Москва, ФИЦ ИУ РАН). **Математические методы анализа данных для обеспечения мониторинга и поддержки предиктивного ремонта элементов жилищной инфраструктуры.**

Данная работа проводится в рамках проекта, нацеленного на разработку комплекса программно-технических решений по созданию информационной системы, обеспечивающей поддержку технического обслуживания элементов жилищной инфраструктуры в целях оперативного контроля их состояния, предиктивного ремонта и оповещения о возникающих нештатных ситуациях на основе технологии Интернета вещей.

Одной из основных составляющих архитектуры разрабатываемой информационной системы [1] является модуль, включающий в себя компоненты предобработки данных, извлечения из данных признаков для последующего применения методов машинного обучения, построения моделей функционирования элементов жилищной инфраструктуры, обнаружения и предсказания нештатных ситуаций. Разработаны и реализованы на языке Python алгоритмы функционирования компонентов модуля.

В рамках компонента *предобработки данных* реализованы следующие методы заполнения пропущенных значений в данных: линейная интерполяция, интерполяция сплайном, простое скользящее среднее, заполнение предыдущим значением. Выполняется очистка данных от шума с использованием фильтра Фурье и нелинейного медианного фильтра. Осуществляется нормализация данных (z -нормализация или min-max нормализация). Разнообразие методов предобработки обусловлено различной природой сигналов с датчиков.

В рамках компонента *извлечения признаков из данных* формируются признаки, относящиеся к дескриптивной статистике (среднее, стандартное отклонение, коэффициенты асимметрии и эксцесса), дифференциальные признаки (первая и вторая производные для временного ряда), спектральные признаки (главный пик спектра), попарные корреляции и свертки; параметры модели ARIMA.

В рамках компонента *построения моделей* реализованы следующие подходы: сезонная интегрированная модель авторегрессии — скользящего среднего SARIMA, скрытые марковские модели (НММ), градиентный бустинг над решающими деревьями (GBM). При *построении модели SARIMA* проводится ограничение множества возможных значений гиперпараметров модели на основании графиков функций автокорреляции и частичной автокорреляции данных, выбираются конкретные значения гиперпараметров, соответствующие минимальному значению информационного критерия Акаике, проверяется статистическая гипотеза о нестационарности остатков временного ряда с использованием расширенного теста Дики-Фуллера. При *построении модели НММ* инициализация параметров модели проводится с помощью обучающей выборки по методу максимума правдоподобия, затем параметры улучшаются с помощью алгоритма Баума-Уэлша. При *построении модели GBM* сначала пространство извлеченных из данных признаков сокращается на основании ранжирования признаков при помощи построения случайного леса. По полученным признакам с использованием кросс-валидации на обучающей выборке и экспертной оценки набора гиперпараметров перебираются варианты моделей, для которых оценивается метрика качества (коэффициент детерминации или среднеквадратичная ошибка).

В рамках компонента *обнаружения нештатных ситуаций* к анализируемым временным рядам применяются построенные модели SARIMA и НММ. Происходит бинарная классификация показаний датчиков на показания, соответствующие штатной работе исправного оборудования, и показания, соответствующие различным нештатным ситуациям (аномалиям во временном ряду). *Предсказание нештатных ситуаций* осуществляется с использованием построенной модели GBM и оценки на ее основании оставшегося времени жизни элемента инфраструктуры.

Реализованные методы были опробованы на готовых наборах данных, состоящих из показаний датчиков бытовых домашних метеостанций (для обнаружения аномалий), а также турбин (для оценки оставшегося времени жизни). Полученные значения метрик качества построенных моделей близки к результатам, приведенным в известных родственных работах.

Работа выполнена при поддержке Минобрнауки России (уникальный идентификатор проекта RFMEFI60717X0176).

СПИСОК ЛИТЕРАТУРЫ

1. *Kovalev D., Shanin I., Stupnikov S., Zakharov V.* Data mining methods and techniques for fault detection and predictive maintenance in housing and utility infrastructure. — In: EnT 2018 — 2018 International Conference on Engineering Technologies and Computer Science. (20–21 March 2018, Moscow.) Proceedings. / Ed. by V. Gvozdetski, N. Pakulin. Piscataway, NJ–Los Alamitos, CA: IEEE, 2018, p. 47–52.