

свободной, затопленной струи. Качественно подобранные входные параметры, сводят отличия лабораторного датчика и виртуального к минимуму. Главный недостаток, отсутствие записи значений датчика давления в таблицу с последующей генерацией графиков зависимости, и отсутствие генерации отчёта с данными для печати.

### Список использованных источников

1. Елисеева Е.В., Злюбина С.Н., Цифровые образовательные ресурсы как составляющая инновационной образовательной среды современного вуза//УДК
2. Баяндин Д.В. Реализация концепции полнофункциональной предметно-ориентированной среды обучения // Образовательные технологии и общество (Educational Technology & Society). 2015. Т. 18. № 4. С. 574–601. [http://ifets.ieee.org/russian/depository/v18\\_i4/pdf/4.pdf](http://ifets.ieee.org/russian/depository/v18_i4/pdf/4.pdf) (дата обращения: 28.04.2021).
3. Лабораторный практикум. [Электронный ресурс]. – URL: <http://media.ls.urfu.ru/460/1178/2484/2458/> (дата обращения: 28.04.2021).

УДК 004.021

**Н. И. Шаханов, И. А. Варфоломеев, Л. Н. Виноградова, О. В. Юдина, Е. В. Ершов**

ФГБОУ ВО «Череповецкий государственный университет», г. Череповец, Россия

## ОБОБЩЕННЫЙ АЛГОРИТМ ФУНКЦИОНИРОВАНИЯ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ОТКАЗОВ ПРОМЫШЛЕННОГО ОБОРУДОВАНИЯ В УСЛОВИЯХ МАЛОГО КОЛИЧЕСТВА ПОЛОМОК

**Аннотация.** Представлен обобщенный алгоритм функционирования системы прогнозирования отказов промышленного оборудования в условиях малого количества поломок. Описаны основные этапы: извлечение данных, обучение моделей с параметрической идентификацией и мониторинг данных для выявления отказов промышленного оборудования.

**Ключевые слова:** алгоритм, прогнозирование отказов, машинное обучение, малое количество неисправностей, промышленное оборудование.

**Abstract.** The report presents a generalized algorithm for the functioning of the system for predicting the failure of industrial equipment in a small number of breakdowns. The main stages are described: data extraction, training of models with parametric identification and data monitoring to detect failures of industrial equipment.

**Key words:** algorithm, failure prediction, machine learning, a small number of malfunctions, industrial equipment.

Эксплуатация промышленного оборудования характеризуется значительными затратами, связанными, в том числе, и с его внеплановыми простоями. Сокращение внеплановых простоев оборудования возможно за счет

повышения эффективности методов планирования, применения автоматизированных систем управления для оценки состояния оборудования и прогнозирования его отказов, что требует разработки новых методов, гарантирующих точность прогноза, в том числе и за счет использования данных, соответствующих реальному состоянию оборудования. Поэтому наиболее перспективным для решения задачи прогнозирования является подход, основанный на машинном обучении.

В настоящее время прогнозирование отказов оборудования на основе методов машинного обучения в условиях малого количества поломок и изменяющихся режимах работы оборудования является затруднительным. В этой связи задача разработки математического и алгоритмического обеспечения для прогнозирования отказов промышленного оборудования является актуальной.

Для реализации метода прогнозирования отказов в условиях малого количества поломок требуется разработать алгоритмического обеспечение, необходимое для дальнейшей программной реализации комплекса прогнозирования отказов промышленного оборудования. Алгоритмическое обеспечение, разработанное применительно к конкретному объекту управления, позволяет определить необходимые структуру и состав вычислительно-управляющего комплекса [1]. Алгоритм представлен в форме структурных блок-схем (рис. 1) с пояснительным текстом и служит основой для разработки программного комплекса прогнозирования отказов промышленного оборудования на основе алгоритмов машинного обучения.

Весь алгоритм можно разбить на три ключевых этапа: извлечение данных, обучение моделей с параметрической идентификацией и мониторинг данных для выявления отказов промышленного оборудования.

На первом этапе проводится извлечение данных из базы данных (БД) с последующей обработкой их в нужном формате для сохранения в распределенную файловую систему Hadoop (HDFS) [2]. Для этого необходимо все доступные данные разбить на три временных отрезка: где отрезки 1 и 2 должны содержать данные только при нормальной работе оборудования. Отрезок 1 должен содержать данные для обучения модели. Отрезок 2 – данные для тестовой выборки (оценка модели и определения средней абсолютной ошибки прогнозирования значения параметра). Отрезок 3 – обязательно требуется наличие зафиксированных отказов. На данных отрезка 3 выполняется параметрическая идентификация.

На втором этапе проводится обучение всех моделей, на основе алгоритма машинного обучения Random Forest [3], для каждого технологического параметра строится своя модель. Далее формируется выходная прогнозная модель, которая группирует все модели Random Forest для каждого технологического параметра и специальную модель ARIMA (методология Бокса–Дженкинса) [4]. Внутри данной модели так же встраивается функционал для определения режима работы оборудования и алгоритмы для работы с матрицами, используются простые математические пакеты.

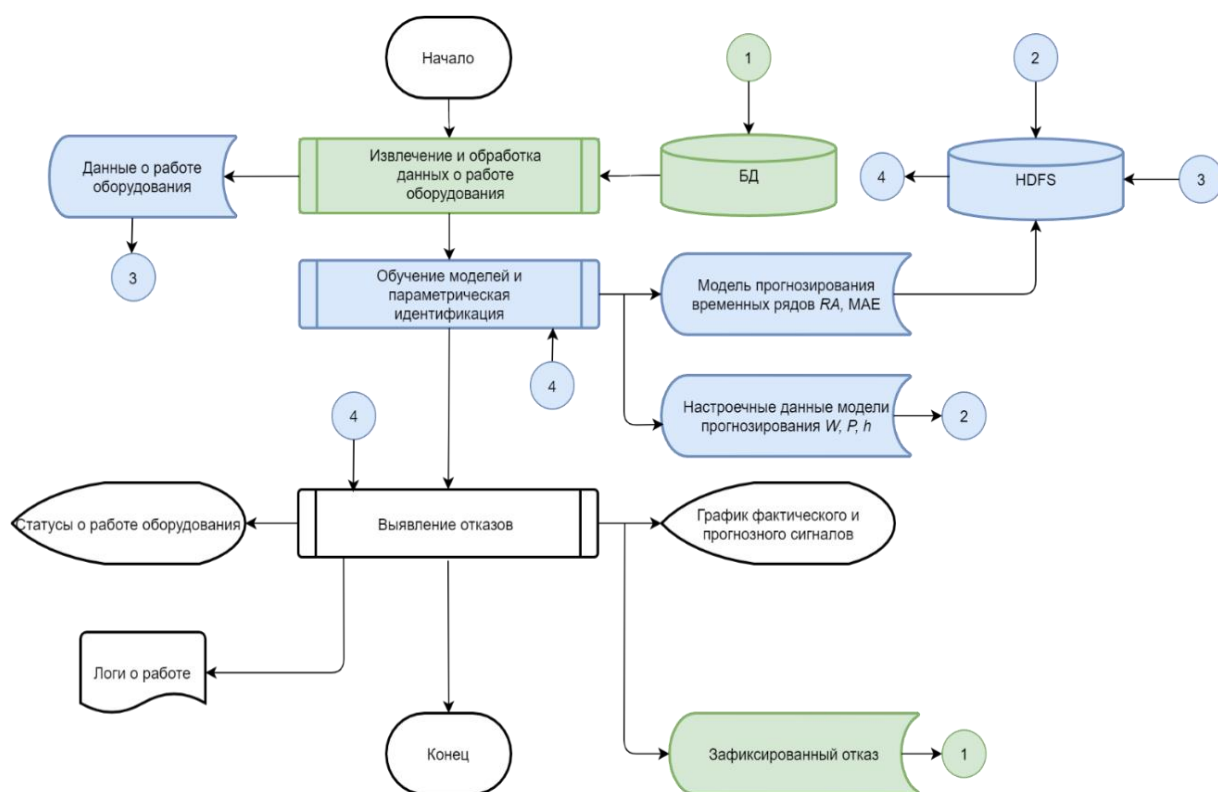


Рис. 1. Блок-схема обобщенного алгоритма функционирования системы прогнозирования отказов промышленного оборудования

Для построения модели прогнозирования временных рядов предполагается использовать стекнинг алгоритмов машинного обучения Random Forest и ARIMA. Обусловлено это тем, что конечная модель дает повышение точности прогнозирования при различных режимах работы оборудования и так же эффективно прогнозирует временные ряды при больших объемах данных.

В качестве скриптового языка прогнозирования используется объектно-ориентированный язык высокого уровня – R. После успешного обучения все модели и итоговая модель сохраняются в HDFS. Скриптовый язык позволяет настраивать этапы обучения модели, корректировать ее параметры, без необходимости разработки специализированного программного обеспечения. В конечном итоге алгоритм обучения модели прогнозирования временных рядов может легко адаптироваться под любое количество входных параметров модели.

Для возможности управлять системой на этапе выявления отказов необходимо задействовать три типа параметров: окно прогнозирования  $W$ , максимальную долю превышения порога  $P$  и порог срабатывания  $h$ , допускающий превышение разности прогнозных и фактических данных в  $k \cdot MAE$  раз для каждого технологического параметра. Для получения прогноза по отказам требуется, чтобы эти параметры были идентифицированы. С их помощью можно управлять моделью прогнозирования отказов, а именно от данных параметров зависит: количество ложных и верных срабатываний, время обработки данных моделью и желаемое время прогнозирования до потенциального отказа. Параметрическая идентификация заключается в поиске

таких значений  $W, h_i, P$  дающие вариант с максимальным результирующим значением формулы 1, при этом найденные значения параметров станут целевыми на этапе выявления отказов. Данные так же отправляются на хранение в HDFS.

$$k_1 = \sum_{i=1}^D \prod_{j=1}^N \begin{cases} 1, & A_j(M_i, M'_i, W, h_j, P) = A'_j(i) \\ 0, & A_j(M_i, M'_i, W, h_j, P) \neq A'_j(i) \end{cases}$$

$$k_2 = \sum_{i=1}^D \prod_{j=1}^N \begin{cases} 0, & A_j(M_i, M'_i, W, h_j, P) = A'_j(i) \\ 1, & A_j(M_i, M'_i, W, h_j, P) \neq A'_j(i) \end{cases} \quad (1)$$

$$C_{\text{верн.}} \cdot k_1 - C_{\text{ложн.}} \cdot k_2 \rightarrow \max,$$

где  $h_i = k_i \cdot MAE_i$ , и каждый коэффициент  $k_i$  входит в диапазон значений  $[0.1, 100]$ ;

$W$  – длина окна прогнозирования, диапазон значений  $[1, +\infty)$ , с;

$P$  – максимальная доля точек в окне, превышающих порог  $h_i$  – не считающийся аномалией в работе  $(0, 1)$ ;

$A_j$  – признак аномалии  $j$ -го параметра;

$A'_j(i)$  – наличие отказа  $j$ -го параметра из  $i$ -й тестовой выборки,  $A'_j = 1$ , когда тестовая выборка взята за несколько часов до зафиксированного отказа  $j$ -го параметра;

$D$  – количество исходных тестовых выборок;

$C_{\text{верн.}}$  – коэффициент при верном прогнозировании отказа(-ов);

$C_{\text{ложн.}}$  – коэффициент при ложных срабатываниях.

Третий этап, используется в режиме реального времени, всего выполняется две итерации, в качестве входных параметров на первой итерации используются фактические данные с контроллера установки и модели прогнозирования временных рядов. На данной итерации проводится прогнозирование всех значений технологических параметров в указанном окне прогнозирования по всем фактическим значениям параметров этого окна. На второй итерации вычисляется наличие аномалии на всех групповых узлах агрегата. Для этого проводится сравнение разности показаний фактических и прогнозных значений всех технологических параметров относительно заданного окна и вычисляется доля превышений этой разности относительно указанного порога. В случае если было обнаружено, что доля отклонений больше указанного при параметрической идентификации, фиксируется аномалия и данные об аномальной работе технологического узла, связанного с этим параметром, и времени выявления отправляются в систему мониторинга работы оборудования, где в последующей перевалке ремонтная группа сможет проверить на наличие неисправностей.

Для возможности проведения экспериментального исследования данное алгоритмическое обеспечение было реализовано в виде системы прогнозирования отказов оборудования уборочной группы прокатного производства, состоящий из: программного обеспечения для получения и сохранения данных с контроллеров агрегата [5]; программного обеспечения прогнозирования отказов оборудования [6]; программного обеспечения для вычислительного комплекса, осуществляющее обучение и параметрическую

идентификацию моделей прогнозирования временных рядов; базы данных, где хранятся актуальные данные с контроллеров; HFDS, где хранятся модели и исторические данные с контроллеров.

*Работа подготовлена при финансовой поддержке Гранта Правительства Вологодской области за 2019 год «Прогнозирование состояния промышленного оборудования на основе анализа данных и машинного обучения»*

### **Список использованных источников**

1. Осколков В.М. Применение параллельных вычислений для прогнозирования на основе алгоритма машинного обучения Random Forest / В.М. Осколков, Н.И. Шаханов, И.А. Варфоломеев, О.В. Юдина, Л.Н. Виноградова, Е.В. Ершов // Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации. Распознавание – 2017 сборник материалов XIII Международной научно-технологической конференции. 2017. С. 267-269.

2. Taylor R. C. An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics // BMC bioinformatics. – BioMed Central, 2010. – Т. 11. – №. 12. – С. S1.

3. Пальмов С.В. Случайный лес: основные особенности / С.В. Пальмов, А.О. Денискова // Наука сегодня: теоретические и практические аспекты. 2017. С. 51-52.

4. Крюков Ю.А. ARIMA – модель прогнозирования значений трафика / Ю.А. Крюков, Д.В. Чернягин // Информационные технологии и вычислительные системы. 2011. № 2. С. 41-49.

5. Шаханов Н.И. Универсальный подход сбора технологической информации с контроллеров промышленных агрегатов / Н.И. Шаханов, В.М. Осколков, И.А. Варфоломеев, О.В. Юдина, Л.Н. Виноградова, Е.В. Ершов // Вузовская наука – региону Материалы XV Всероссийской научной конференции с международным участием. 2017. С. 113-115.

6. Шаханов Н.И. Прогнозирование аномалий в работе натяжных устройств агрегата полимерных покрытий металла в условиях малого количества поломок / Н.И. Шаханов, И.А. Варфоломеев, Е.В. Ершов, О.В. Юдина // Программные продукты и системы. 2018. С. 2012-2018.