

DOI 10.23859/1994-0637-2020-4-97-3

УДК: 004.8

© Ершов Е. В., Юдина О. В.,
Виноградова Л. Н., Шаханов Н. И., 2020

Ершов Евгений Валентинович

Доктор технических наук, профессор,
Череповецкий государственный университет
(Череповец, Россия)
ORCID 0000-0003-2888-4242
E-mail: evershov@chsu.ru

Ershov Evgeny Valentinovich

Doctor of Technology, Professor,
Cherepovets State University
(Cherepovets, Russia)
ORCID 0000-0003-2888-4242
E-mail: evershov@chsu.ru

Юдина Ольга Вадимовна

Кандидат технических наук,
Череповецкий государственный университет
(Череповец, Россия)
E-mail: ovyudina2010@mail.ru

Yudina Olga Vadimovna

PhD in Technology,
Cherepovets State University
(Cherepovets, Russia)
E-mail: ovyudina2010@mail.ru

Виноградова Людмила Николаевна

Кандидат технических наук,
Череповецкий государственный университет
(Череповец, Россия)
E-mail: LNVinogradova@bk.ru

Vinogradova Lyudmila Nikolaevna

PhD in Technology,
Cherepovets State University
(Cherepovets, Russia)
E-mail: LNVinogradova@bk.ru

Шаханов Никита Иванович

Кандидат технических наук,
Череповецкий государственный университет
(Череповец, Россия)
E-mail: nikshakhanov@gmail.com

Shakhanov Nikita Ivanovich

PhD in Technology,
Cherepovets State University
(Cherepovets, Russia)
E-mail: nikshakhanov@gmail.com

**МОДЕЛИРОВАНИЕ СОСТОЯНИЯ
ОБОРУДОВАНИЯ НА ОСНОВЕ
СТЕКИНГА АЛГОРИТМОВ
МАШИННОГО ОБУЧЕНИЯ RANDOM
FOREST И ARIMA**

**EQUIPMENT CONDITION MODELING
BASED ON RANDOM FOREST AND
ARIMA MACHINE LEARNING
ALGORITHM STACKING**

Аннотация. В статье рассматриваются алгоритмы построения прогнозной модели состояния промышленного оборудования с использованием методов анализа данных и машинного обучения. В основу построения модели положены алгоритмы Random Forest (RF) и ARIMA (AR), изучены особенности применения алгоритмов, обучения моделей и подбора оптимальных параметров; показана необходимость использования второго алгоритма. Приведена структурная схема модели прогнозирования временных рядов с использованием стекинга, оценка результатов моделирования.

Abstract. The article discusses algorithms for constructing predicative models of the industrial equipment condition using data analysis and machine learning. The model is based on Random Forest (RF) and ARIMA (AR) algorithms. The authors consider approaches to learning algorithms and optimizing parameters. A block diagram of a time series predictive model applying stacking is presented, as well as an assessment of the simulation results.

Ключевые слова: прогноз состояния оборудования, режимы работы, машинное обучение, стекинг алгоритмов

Keywords: equipment condition forecast, the operation modes of the equipment, machine learning, algorithm stacking

Благодарность. Статья подготовлена при финансовой поддержке гранта Правительства Вологодской области за 2019 год «Прогнозирование состояния промышленного оборудования на основе анализа данных и машинного обучения».

Введение

Ремонты и техническое обслуживание промышленного оборудования являются одной из основных статей расходов предприятия, росту этих затрат способствуют традиционные методы планирования ремонтов – предупредительные, ремонт после отказа. Для того чтобы иметь возможность использовать современные стратегии управления ремонтами, предприятию необходим прогноз отказов оборудования, созданный на основе данных как прошедших периодов, так и текущего состояния устройств. Получить такие прогнозы позволяют методы предиктивной аналитики, базирующиеся на анализе данных и машинном обучении. Рассматриваемые алгоритмы разрабатывались для прогнозирования отказов оборудования уборочной группы прокатного производства и прошли экспериментальную проверку в листо-прокатном цехе № 2 производства горячекатаного проката ПАО «Северсталь».

Для получения прогноза состояния оборудования с помощью технологии машинного обучения могут быть использованы различные подходы, связанные с построением регрессий, зависимостей других типов, деревьев, в частности деревьев решений, кластеризации. Для оборудования уборочной группы прокатного производства определено, что прогноз состояния устройств будет выполнен на основе анализа временных рядов крутящих моментов двигателей роликов отводящего рольганга. В основу моделирования положен подход, предполагающий прогнозирование оптимальной работы оборудования¹, поэтому используются данные, соответствующие нормальной работе двигателей роликов отводящего рольганга.

Основная часть

В результате исследования², принимающего во внимание требования к способности алгоритма учитывать необходимое количество измерений, скорости вычислений и точности результатов, был выбран алгоритм Random Forest (RF)³, основанный на построении решающих деревьев.

¹ Шаханов Н. И., Юдина О. В., Ершов Е. В., Виноградова Л. Н., Мишутушкин М. А., Варфоломеев И. А. Оценка состояния отводящего рольганга стана 2000 ПАО «Северсталь» на основе анализа данных и машинного обучения // Вестник Череповецкого государственного университета. – 2020. – № 2 (95). – С. 47–55.

² Шаханов Н. И., Варфоломеев И. А., Ершов Е. В., Юдина О. В. Прогнозирование отказов роликов отводящего рольганга при производстве горячекатаного проката // Производство проката. – 2018. – № 7. – С. 9–14.

³ Летова М. С. Реализация регрессивных и классификационных задач с помощью метода Random Forest // E-Scio. – 2017. – № 8 (11). – С. 15–21; Svetnik V., Liaw A., Tong C., Culbertson J. C., Sheridan R. P., Feuston B. P. Random Forest: A Classification and Regression Tool

Для оценки модели прогнозирования временных рядов крутящих моментов приводных электродвигателей роликов в качестве входных использованы данные, накопленные в существующей автоматизированной системе управления технологическими процессами (далее – АСУТП).

Для обучения, оценки и корректировки модели входные данные были разделены на 3 основных отрезка (см. рис. 1):

Отрезок 1 – обучение модели.

Отрезок 2 – оценка модели и вычисление средней абсолютной ошибки (MAE).

Отрезок 3 – вычисление коэффициента детерминации R^2 и оценки значимости t -критерия Стьюдента.

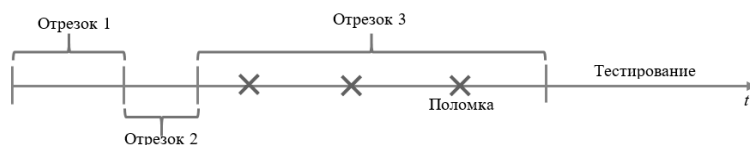


Рис. 1. Разделение обучающей выборки

Оценка модели RF производилась на основе вычисленной средней абсолютной ошибки MAE для каждого ролика отводящего рольганга:

$$MAE_i = \frac{1}{K} \sum_{t=1}^K |M_i(t) - M'_i(t)|, \quad (1)$$

где $i \in 1 \dots N$, K – количество точек из тестовой выборки; M'_i – прогнозное значение крутящего момента в момент времени t , M_i – фактическое значение крутящего момента в момент времени t .

Обучение модели по алгоритму Random Forest включает этапы подготовки данных, определения параметров и оценки модели. Этап подготовки данных для обучения на отрезке 1 в нашем случае предполагает ликвидацию пропусков информации и исключение из рассмотрения последовательностей, относящихся к другим режимам работы. Пропуски возможны в случае неполных данных на том или ином интервале времени по техническим причинам – запись о таком временном промежутке удаляется целиком. Ранее отмечалось¹, что модель использует сведения только о нормальной работе оборудования, поэтому на этапе подготовки также удаляются записи о зафиксированных отказах в работе роликов или работе любого из роликов в аномальном режиме.

Использование алгоритма Random Forest при построении прогнозной модели предполагает настройку основных параметров алгоритма: допустимого количества

for Compound Classification and QSAR Modeling // Journal of Chemical Information and Computer Sciences. – 2003. – Vol. 43. – № 6. – P. 1947–1958.

¹ Шаханов Н. И., Юдина О. В., Ершов Е. В., Виноградова Л. Н., Миштушкин М. А., Варфоломеев И. А. Оценка состояния отводящего рольганга стана 2000 ПАО «Северсталь» на основе анализа данных и машинного обучения // Вестник Череповецкого государственного университета. – 2020. – № 2 (95). – С. 47–55.

деревьев решений (или глубины деревьев) для описания данных и максимального допустимого количества точек в каждом листе дерева. Последний параметр напрямую влияет на точность прогнозируемого значения, первый определяет точность моделирования данных в целом: чем больше глубина, тем точность моделирования выше, но он же и обратно пропорционален производительности модели при вычислении временных рядов крутящих моментов. Избыточное допустимое число деревьев может снизить производительность модели и не позволит получить прогноз в реальном времени. Точность модели оценивается по средней абсолютной ошибке MAE, точность прогнозного значения – по коэффициенту R^2 .

Результаты обучения алгоритма представлены в табл. 1.

Таблица 1

Влияние глубины деревьев алгоритма RF на абсолютную ошибку модели прогнозирования временных рядов

Количество объектов в листе	Глубина деревьев				
	50	25	10	5	2
1000	0,58039	0,58298	0,60493	0,63833	0,6899
500	0,43949	0,44871	0,49515	0,48274	0,5760
400	0,41744	0,42057	0,43453	0,45978	0,5314
300	0,37873	0,38562	0,40287	0,41992	0,48851
200	0,33667	0,34171	0,35454	0,38135	0,46452
100	0,29137	0,29569	0,31002	0,33174	0,38409
50	0,25883	0,26543	0,27807	0,29645	0,35202
22	0,23657	0,24317	0,25366	0,27945	0,33343
5	0,22324	0,22922	0,24801	0,26699	0,32708

Тестирование модели производилось с учетом результатов обучения, они же использовались и для определения оптимальных значений параметров алгоритма. На их основе была построена модель прогнозирования временных рядов крутящих моментов двигателей. Исследование показало, что в случае, когда количество точек узла меньше 22, модель перестает обрабатывать входной массив данных в режиме реального времени при большом его объеме.

В результате тестирования было определено оптимальное количество деревьев решений – 50 и допустимое количество точек в листе – 22 при $MAE = 0,23657$. Следовательно, модель с 50 деревьями и максимальным количеством точек в разбиении до 22 дает объяснение в среднем $99 \pm 0,7$ % данных в выборке для обучения в зависимости от типа двигателя и $98,2$ % данных в тестовой выборке.

Однако с помощью одного алгоритма оказалось невозможным получить модели с требуемыми характеристиками для режима холостого хода рольганга.

Было установлено, что в режиме работы электродвигателей без нагрузки (происходит вращение роликов, но листа нет) в течение длительного времени модель Random Forest не позволяет получить прогноз крутящего момента требуемой точности по каждому параметру.

Принято считать, что решением в подобных случаях может быть увеличение обучающей выборки, это позволит учесть все основные режимы работы оборудова-

ния. Однако увеличение входной выборки в два раза позволило получить лишь незначительный прирост точности модели при холостом режиме работы оборудования – в пределах половины процента, но при этом наблюдалось заметное ухудшение функционирования модели в остальных режимах работы.

Кроме того, чрезмерное увеличение обучающей выборки вызывает одну из проблем машинного обучения – переобучение модели, проиллюстрированное для рассматриваемого алгоритма на рис. 2. Оно сводится к подстройке параметров модели под параметры обучающей выборки и ее фактическому копированию, что приводит к снижению качества прогнозирования на реальных данных.

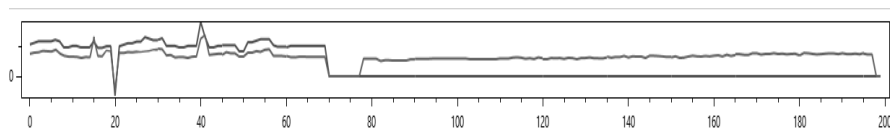


Рис. 2. Подстройка параметров переобученной моделью

Для обработки данных по крутящим моментам приводных электродвигателей и создания прогноза работы оборудования в указанные моменты времени необходим другой алгоритм. Наиболее подходящим оказался алгоритм ARIMA¹, использующий интегрированные модели авторегрессии и скользящего среднего, что позволяет успешно применять его для прогнозирования в том числе и нестационарных временных рядов.

Поскольку для построения прогнозной модели по полному набору данных необходимы оба алгоритма, был применен стекинг используемых алгоритмов машинного обучения.

Для улучшения качества прогнозирования крутящего момента при всех режимах работы и исключения проблемы переобучения принято решение использовать стекинг двух алгоритмов. Стандартная модель стекинга может быть описана формулой и предполагает возможность переключения на другую модель для холостого хода (2). Введем в стандартную модель изменение, которое позволит в дальнейшем автоматически выбрать модель прогнозирования в зависимости от режима работы оборудования (3):

$$K = 2,$$

$$w_1 = \begin{cases} 1, & \text{если холостой ход} \\ 0, & \text{иначе} \end{cases},$$

$$w_2 = \begin{cases} 0, & \text{если холостой ход} \\ 1, & \text{иначе} \end{cases},$$

(2)

¹ Крюков Ю. А., Чернягин Д. В. ARIMA – модель прогнозирования значений трафика // Информационные технологии и вычислительные системы. – 2011. – № 2. – С. 41–49; Пилугина А. В., Бойко А. А. Использование моделей ARIMA для прогнозирования валютного курса // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 4 (32). – С. 249–267.

$$f(x) = \sum_{k=1}^K w_k f_k(x), \quad (3)$$

где K – количество моделей; $f_k(x)$ – алгоритм регрессии; w – вес модели.

Для режима холостого хода была выбрана интегрированная авторегрессионная модель со скользящим средним ARIMA (AR) (4), которая, в отличие от Random Forest, использует для моделирования работы каждого электродвигателя соответствующие ему данные одного крутящего момента и определяет количество входных точек в соответствии с размером окна прогнозирования.

Для прогнозирования значений временных рядов крутящих моментов модель AR вычисляет значения временных рядов следующим образом:

$$AR(M) = ARIMA(p, d, q), \quad (4)$$

где p, d, q – параметры модели, подбираемые в автоматическом режиме: p – порядок авторегрессионной части модели; d – количество последовательных разностей, взятых для временного ряда $M[k]$; q – порядок части скользящего среднего.

$M[k]$ – значения временного ряда крутящего момента электродвигателя, взятые в k -тые моменты времени.

Модель ARIMA(p, d, q) в общем виде для $M[k]$ представлена следующим образом:

$$M[k] = \varphi_0 + \varphi_1 M[k-1] + \varphi_2 M[k-2] + \dots + \varphi_p M[k-p] + \varepsilon_k - \omega_1 \varepsilon_{k-1} + \omega_2 \varepsilon_{k-2} - \dots - \omega_q \varepsilon_{k-q}, \quad (5)$$

где φ_i – коэффициенты авторегрессионной части модели (оцениваемые параметры); ω_i – коэффициенты скользящего среднего (оцениваемые параметры); ε_i – случайное возмущение, описывающее влияние переменных, не учтенных в модели (5).

Входная M и выходная M' матрицы модели с учетом использования алгоритма ARIMA изменятся и будут иметь вид:

$$M = \begin{bmatrix} M_{1-2W-1} & \dots & M_{1_t} \\ \vdots & \ddots & \vdots \\ M_{N_t-2W-1} & \dots & M_{N_t} \end{bmatrix}, \quad (6)$$

$$M' = \begin{bmatrix} M'_{1_t-W} & \dots & M'_{1_t} \\ \vdots & \ddots & \vdots \\ M'_{N_t-W} & \dots & M'_{N_t} \end{bmatrix}.$$

Увеличение длины временного ряда входных параметров обусловлено тем, что авторегрессионная модель со скользящим средним требует для вычислений временного ряда большей продолжительности¹.

Результирующая модель прогнозирования значений временного ряда будет иметь вид:

$$M' = RA(M) = \begin{cases} AR(M), & \text{для холостого хода} \\ RF(M), & \text{для других режимов} \end{cases} \quad (7)$$

Структурная схема модели прогнозирования, включающей взаимодействие двух алгоритмов, представлена на рис. 3.

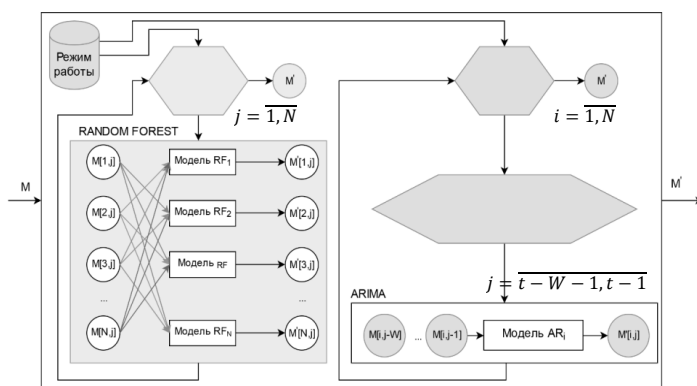


Рис. 3. Структурная схема модели прогнозирования временных рядов

Модель прогнозирования временных рядов крутящих моментов с применением стекинга двух алгоритмов для определения режимов работы агрегата использует значение признака режима холостого хода. Это значение уже было получено в имеющейся АСУТП и было доступно для расчетов при тестировании модели. В зависимости от значения признака выбирается алгоритм прогнозирования.

Тестирование стекинга алгоритмов показало высокую точность прогнозирования на всех режимах работы оборудования (см. рис. 4).

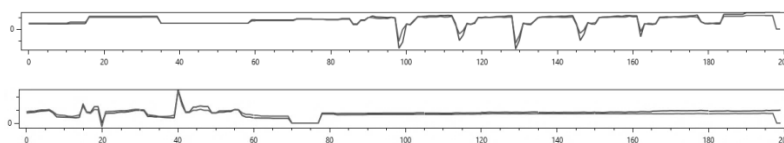


Рис. 4. Графики прогнозного и фактического значения крутящего момента для 1-го и 63-го ролика соответственно с применением модели стекинга алгоритмов RF и AR

В табл. 2 представлены значения средней абсолютной ошибки, коэффициента детерминации, значимость которых проверена по *t*-критерию Стьюдента (значения

¹ Шаханов Н. И., Варфоломеев И. А., Ершов Е. В., Юдина О. В. Прогнозирование отказов роликов отводящего рольганга при производстве горячекатаного проката // Производство проката. – 2018. – № 7. – С. 9–14.

усреднены по всем электродвигателям) при использовании стекинга алгоритмов Random Forest и ARIMA и без него.

Таблица 2

Оценка точности модели с использованием стекинга алгоритмов машинного обучения Random Forest и ARIMA

Используемая модель	MAE	R^2
RF	0,23	0,995
RF+ARIMA	0,20	0,997

Выводы

Мы видим, что применение стекинга алгоритмов позволило повысить точность моделирования и улучшить результат прогноза. Таким образом, модель прогнозирования временных рядов крутящих моментов учитывает режимы работы оборудования при больших объемах данных и позволяет получить прогноз состояния оборудования с требуемой точностью.

Литература

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

Летова М. С. Реализация регрессивных и классификационных задач с помощью метода Random Forest // E-Scio. – 2017. – № 8 (11). – С. 15–21.

Пилюгина А. В., Бойко А. А. Использование моделей ARIMA для прогнозирования валютного курса // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 4 (32). – С. 249–267.

Трегуб А. В. Методика построения модели ARIMA для прогнозирования динамики временных рядов // Лесной вестник. – 2011. – № 5. – С. 179–183.

Шаханов Н. И., Варфоломеев И. А., Ершов Е. В., Юдина О. В. Прогнозирование отказов роликов отводящего рольганга при производстве горячекатаного проката // Производство проката. – 2018. – № 7. – С. 9–14.

Шаханов Н. И., Юдина О. В., Ершов Е. В., Виноградова Л. Н., Мишутушкин М. А., Варфоломеев И. А. Оценка состояния отводящего рольганга стана 2000 ПАО «Северсталь» на основе анализа данных и машинного обучения // Вестник Череповецкого государственного университета. – 2020. – № 2 (95). – С. 47–55.

Svetnik V., Liaw A., Tong C., Culberson J. C., Sheridan R. P., Feuston B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling // Journal of Chemical Information and Computer Sciences. – 2003. – Vol. 43. – № 6. – P. 1947–1958.

References

Kriukov Iu. A., Cherniagin D. V. ARIMA – model' prognozirovaniia znachenii trafika [An ARIMA model for forecasting values of network traffic]. *Informatsionnye tekhnologii i vychislitel'nye sistemy* [Journal of information technologies and computing systems], 2011, no. 2, pp. 41–49.

Letova M. S. Realizatsiia regressivnykh i klassifikatsionnykh zadach s pomoshch'iu metoda Random Forest [Solving regression and classification problems applying Random Forest algorithm]. *E-Scio* [E-Scio], 2017, no. 8 (11), pp. 15–21.

Piliugina A. V., Boiko A. A. Ispol'zovanie modelei ARIMA dlia prognozirovaniia valiutnogo kursa [Using ARIMA models for forecasting currency exchange rate]. *Prikaspiiskii zhurnal: upravlenie i vysokie tekhnologii* [Caspian journal: management and high technologies], 2015, no. 4 (32), pp. 249–267.

Tregub A. V. Metodika postroeniia modeli ARIMA dlia prognozirovaniia dinamiki vremennykh riadov [The procedure of construction the ARIMA model for forecasting the time series dynamics]. *Lesnoi vestnik* [Forestry bulletin], 2011, no. 5, pp. 179–183.

Shakhanov N. I., Varfolomeev I. A., Ershov E. V., Iudina O. V. Prognozirovanie otkazov rolikov otvodiashchego rol'ganga pri proizvodstve goriachekatanogo prokata [Failure prediction of rolls in the collecting roller table when producing hot-rolled products]. *Proizvodstvo prokata* [Rolled products manufacturing], 2018, no. 7, pp. 9–14.

Shakhanov N. I., Iudina O. V., Ershov E. V., Vinogradova L. N., Mishutushkin M. A., Varfolomeev I. A. Otsenka sostoiianiia otvodiashchego rol'ganga stana 2000 PAO “Severstal” na osnove analiza dannykh i mashinnogo obucheniia [Assessing the condition of the discharge roller table at the rolling mill 2000 PAO “SEVERSTAL” based on data analysis and machine learning]. *Vestnik Cherepovetskogo gosudarstvennogo universiteta* [Bulletin of the Cherepovets State University], 2020, no. 2 (95), pp. 47–55.

Svetnik V., Liaw A., Tong C., Culberson J. C., Sheridan R. P., Feuston B. P. Random Forest: a Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences*, 2003, vol. 43, no. 6, pp. 1947–1958.

Для цитирования: Еришов Е. В., Юдина О. В., Виноградова Л. Н., Шаханов Н. И. Моделирование состояния оборудования на основе стекинга алгоритмов машинного обучения Random Forest и ARIMA // Вестник Череповецкого государственного университета. – 2020. – № 4 (97). – С. 32–40. DOI: 10.23859/1994-0637-2020-4-97-3

For citation: Ershov E. V., Yudina O. V., Vinogradova L. N., Shakhanov N. I. Equipment condition modeling based on Random Forest and ARIMA machine learning algorithm stacking. *Bulletin of the Cherepovets State University*, 2020, no. 4 (97), pp. 32–40. DOI: 10.23859/1994-0637-2020-4-97-3