

ИНФОРМАЦИОННОЕ, МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ТЕХНИЧЕСКИХ СИСТЕМ

УДК 621.31

[https://doi.org/10.18503/2311-8318-2021-3\(52\)-57-65](https://doi.org/10.18503/2311-8318-2021-3(52)-57-65)

Антоненков Д.В., Матренин П.В.

Новосибирский государственный технический университет

ИССЛЕДОВАНИЕ АНСАМБЛЕВЫХ И НЕЙРОСЕТЕВЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧЕ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ ГОРНЫХ ПРЕДПРИЯТИЙ

В статье рассмотрена проблема прогнозирования электропотребления горных предприятий, особенностью которого является высокий уровень нестационарности и стохастичности. Используются собранные авторами данные за четыре года по горному предприятию Якутии, работающему в области добычи и переработки угля. При этом отдельно выполнен анализ по различным объектам предприятия: угольному разрезу и обогащательным фабрикам, имеющим принципиально различные технологические процессы, и, следовательно, графики электропотребления. Проведено исследование двух классов методов машинного обучения: обработка ретроспективных данных электропотребления предприятия как временного ряда с помощью рекуррентных нейронных сетей; выделение наиболее значимых признаков для применения к ним ансамблевых моделей на базе деревьев решений: случайного леса, адаптивного бустинга и экстремального градиентного бустинга. Поскольку для указанных моделей машинного обучения очень важна настройка гиперпараметров, для корректного сопоставления результатов проведена процедура оптимизации гиперпараметров всех моделей. Проведенные вычислительные эксперименты показали, что рекуррентные многослойные нейронные сети способны использовать для прогнозирования временные ряды без предварительной обработки, обучаясь выделять значимые признаки из динамики изменения графика электропотребления. Для применения ансамблей регрессионных деревьев решений необходим предварительный анализ данных для отбора из временного ряда наиболее значимых признаков. На примере рассмотренного предприятия показано, что использование такого подхода при работе с ансамблевыми моделями дает близкую точность к рекуррентным нейронным сетям. При этом ансамблевые модели обучаются на 1–2 порядка быстрее, а недостатком является большая склонность к переобучению.

Ключевые слова: горное предприятие, прогнозирование электропотребления, машинное обучение, рекуррентные нейронные сети, выбор признаков, ансамблевые методы.

ВВЕДЕНИЕ

Прогнозирование электропотребления необходимо для экономически эффективной работы всей электроэнергетической системы и для повышения ее устойчивости. Чем выше точность прогнозов от потребителей электроэнергии, тем лучше поставщики способны планировать генерацию и распределение электроэнергии, что в конечном счете снижает издержки всех участников рынка электроэнергии [1].

В настоящее время существует несколько наиболее распространенных методов прогнозирования электропотребления, то есть графиков нагрузки. Можно выделить большую группу методов, основанных на статистических моделях [2, 3], включая сезонные модели Хольта-Уинтерса [4, 5], авторегрессию и различные гибридные методы, такие как ARIMA [5–7], методы параметрического синтеза предопределенных моделей временного ряда [8], соединение авторегрессионных моделей с аппаратом нечеткой логики [9]. Недостатком указанных методов является снижение точности в случаях нестационарного стохастического временного ряда, в котором недостаточно выделить тренд и периодические составляющие. Существующие методы прогнозирования, учитывающие техноэкономические свойства горного предприятия, сводятся к понятию устойчивых рангов гиперболических распределений [10].

Другим направлением исследований является применение методов машинного обучения, таких как ме-

тод опорных векторов [11–13], метод ближайших соседей [14], деревья решений [15], ансамблевые методы [13, 16], искусственные нейронные сети (ИНС). ИНС имеют большое число базовых архитектур, для прогнозирования электропотребления успешно применяются сверточные сети [17, 18], глубокие сети [19–21], чаще других используются рекуррентные сети [18, 21–23] (Recurrent neural networks, RNN), поскольку они созданы для обработки последовательных данных, в частности временных рядов.

Сравнение глубоких нейронных сетей с классическим многослойным перцептроном, ARIMA и сезонной моделью Хольта-Уинтерса для прогноза электропотребления промышленных предприятий приведено в работе [19]. Недостатками глубоких нейронных сетей являются необходимость обучения на очень большой выборке данных, высокая вычислительная трудоемкость обучения [20], кроме того, высокая трудоемкость выбора архитектуры сети [19, 23]. При этом существует проблема переиспользования обученной сети. Например, если для прогноза электропотребления модель в качестве входа использовала метеорологические данные (температура, скорость ветра, давление, облачность), то без них модель уже не будет работать. И может намного хуже работать в другой климатической зоне. Кроме того, модель, обученная на данных определенных промышленных предприятий, может показывать низкую точность прогнозов для предприятия с особым режимом работы или даже аналогичного предприятия в другой стране из-за административных факторов. Все это может приводить к тому, что резуль-

таты нейросетевой модели при неверной настройке окажутся даже хуже, чем результаты простейших методов прогнозирования.

Горное предприятие представляет собой сложный электротехнический комплекс и, в отличие от многих других промышленных объектов, является не только техническим, но и природно-техническим комплексом. Также, в отличие от других промышленных предприятий, горное предприятие динамично развивается не только во времени, но и в пространстве [10, 23]. Это связано с тем, что основные потребители (буровые станки и экскаваторы) перемещаются по карьеру, и сам карьер тоже непрерывно меняет свою поверхность в ходе добычи угля.

МЕТОДЫ ИССЛЕДОВАНИЯ

Постановка задачи

Задача прогнозирования сформулирована следующим образом. Необходимо построить модель, выполняющую преобразование входных данных $\{X, T\}$ в прогноз электропотребления на сутки (24–48 часов) вперед Y^* :

$$Y^* = f(\{X, T\}); \quad (1)$$

$$X = \{x_1, x_2, \dots, x_{24m}\}; \quad (2)$$

$$T = \{t_1, t_2, \dots, t_5\}; \quad (3)$$

$$Y^* = \{y_1^*, y_2^*, \dots, y_{24}^*\}, \quad (4)$$

где x_1, x_2, \dots, x_{24m} – ретроспективные данные почасового электропотребления за m предыдущих суток; t_1, t_2, \dots, t_5 – час суток, день недели, число, месяц и год того часа, на который приходится начало прогноза; $y_1^*, y_2^*, \dots, y_{24}^*$ – почасовой прогноз электропотребления на сутки.

На **рис. 1** приведена графическая интерпретация векторов X и Y^* при $m = 2$. Показателем точности прогноза в данной работе выбран средний модуль процента ошибки (MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i^* - y_i}{y_i} \right| \cdot 100\%, \quad (5)$$

где n – число часов в выборке; y_i^* – прогноз электропотребления в i -й час; y_i – истинное значение электропотребления в i -й час.

Выбор MAPE обусловлен тем, что именно этот показатель используется предприятиями как главная метрика точности прогноза электропотребления при работе на оптовом и розничном рынках электроэнергии и мощности.

Сутки	1:00	2:00	3:00	...	0:00
2 суток назад	x_1	x_2	x_3	...	x_{24}
1 сутки назад	x_{25}	x_{26}	x_{27}	...	x_{48}
Текущие сутки					
Прогноз на сутки вперед	y_1^*	y_2^*	y_3^*	...	y_{24}^*

Рис. 1. Векторы X и Y^* из выражений (1), (2) и (4)

Рекуррентная нейросетевая модель

В данной работе среди множества архитектур ИНС выбрана рекуррентная, как наиболее подходящая для обработки временных рядов (вектор X из выражений (1), (2)). Число слоев и другие архитектурные гиперпараметры сети подбирались экспериментально, в соответствии с процедурой, описанной в [23]. Для задач обработки временных рядов целесообразно применять рекуррентные сети, которые обладают свойством долгосрочной памяти. Более эффективно управляют своей памятью рекуррентные сети, основанные на ячейках Long short-term memory (LSTM) [24] и Gated Recurrent Unit (GRU) [25]. В исследовании использованы ячейки обоих типов, наилучший результат получен с использованием GRU.

В качестве пороговой функции скрытых полносвязных слоев использована ReLU [26].

Для борьбы с переобучением использованы приемы Dropout [23, 27], отключающий в случайном порядке отдельные нейроны сети и L2-регуляризация [23, 28], которая добавляет функции потерь обучения сумму квадратов весов сети.

Среди множества методов обучения ИНС следует выделить метод обратного распространения ошибки (Backpropagation), именно он чаще всего используется в обучении ИНС. Метод обратного распространения ошибки имеет большое число модификаций. В данном исследовании использован метод Adam [29]:

$$\begin{aligned} V_{dw} &= \beta_1 V_{dw} + (1 - \beta_1) dW; \\ S_{dw} &= \beta_2 S_{dw} + (1 - \beta_2) dW^2; \\ V_{dw}^{corr} &= V_{dw} (1 - \beta_1^t); \\ S_{dw}^{corr} &= S_{dw} (1 - \beta_2^t); \\ W &= W - \alpha V_{dw}^{corr} \left(\sqrt{S_{dw}^{corr} + \varepsilon} \right)^{-1}, \end{aligned} \quad (6)$$

где V_{dw} – матрица, характеризующая инерционные свойства параметров ИНС, по сути, матрица скорости изменения параметров; β_1 – параметр, задающий баланс между учетом предыдущего направления градиента и направления градиента, полученного на очередной эпохе обучения и на очередном пакете, обычно значение этого параметра близко к 1 (~0,9); dW – матрица градиентов, задающих направление повышения ошибки, $\partial E / \partial W$; β_2 – параметр, задающий баланс между учетом предыдущей энергии изменения направления градиента и направления градиента, полученного на очередной эпохе обучения и на очередном пакете, обычно значение этого параметра близко к 1 (~0,999); S_{dw} – матрица, характеризующая степень («энергию», так как градиент возведен в квадрат) изменения параметров ИНС, без учета направления изменения; t – номер пакета при обучении; W – матрица весов; α – величина шага обучения; ε – близкое к нулю положительное число для предотвращения деления на ноль.

Архитектура используемой сети показана на **рис. 2** (GRU – gate recurrent unit, рекуррентный слой, FC – fully connected, полносвязный слой, sigmoid – сигмоидальная функция активации).

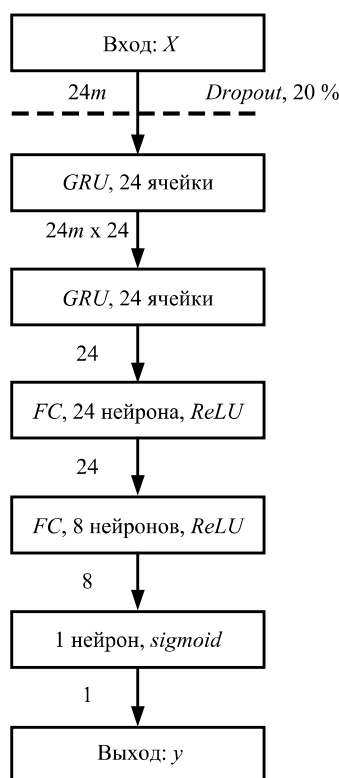


Рис. 2. Архитектура используемой рекуррентной нейронной сети

Отбор признаков

Рекуррентные нейронные сети позволяют обрабатывать последовательные данные, то есть обрабатывать целиком временной ряд фактического почасового электропотребления предприятия за прошлые m суток. Это позволяет выделять сложные зависимости между ретроспективными данными и будущим электропотреблением. В то же время сложный технологический процесс горных предприятий может быть таковым, что из всего графика электропотребления за прошлые сутки не следуют изменения, которые будут присутствовать в следующих сутках. Поэтому целесообразно провести анализ признаков и выделить наиболее значимые из них. Это позволит применить для решения задачи методы машинного обучения, не требующие обработки временного ряда, во-первых, для сравнения результатов рекуррентных ИНС с принципиально другими подходами, во-вторых, кратное снижение числа признаков в разы повышает скорость обучения моделей машинного обучения, в-третьих, меньшее число параметров может повысить точность обучения за счет меньшего риска обнаружения ложных зависимостей.

Поэтому в ходе описанных далее экспериментальных исследований в работе из $\{X, T\}$ отбирались наиболее значимые признаки по критериям Пирсона и Спирмена.

Ансамблевые методы машинного обучения

Ансамблевые методы машинного обучения основаны на системном эффекте, который возникает при объединении в одну модель множества отдельных простых моделей. При этом точность полученной системы намного выше точностей ее элементов.

В работе рассмотрены три подхода к построению ансамблевых моделей: случайный лес (Random Forest [30, 31]), адаптивный бустинг (AdaBoost [31, 32]) и градиентный бустинг (XGBoost [33]).

Алгоритм случайного леса можно сформулировать следующим образом:

- 1) разделить обучающую выборку данных на s случайных подвыборок методом Монте-Карло, один и тот же элемент может попадать в разные подвыборки;
- 2) для каждой подвыборки построить регрессионную модель на базе дерева решений b_i , $i = 1, \dots, s$;
- 3) итоговая модель для рассматриваемой задачи:

$$Y^*(\{X^*\}) = \frac{1}{s} \sum_{i=1}^s b_i(X^*). \quad (7)$$

В выражении (7) используется обозначение X^* , чтобы показать, что в данной работе применение метода следует после отбора наиболее значимых признаков.

В случае применения случайного леса каждый отдельный регрессор строится независимо от результатов остальных, так что они могут строиться параллельно. Бустинг принципиально отличается от случайного леса тем, что регрессоры строятся последовательно и каждый новый зависит от текущих результатов предыдущих регрессоров, объединенных в модель. Кроме того, используется взвешенная сумма результатов отдельных регрессоров:

$$Y^*(\{X^*\}) = \sum_{i=1}^s \gamma_i h_i(X^*), \quad (8)$$

где γ_i – вес регрессора; h_i – регрессор.

Главным отличием AdaBoost и XGBoost друг от друга является способ учета ошибок ансамбля. В адаптивном бустинге на каждой итерации увеличивается вес объектов обучающей выборки, на которых была допущена ошибка. В градиентном бустинге каждый последующий регрессор строится так, чтобы обеспечить максимальное снижение ошибки ансамбля, градиент определяет направление наискорейшего убывания ошибки.

Вычислительные эксперименты и полученные результаты

Выборка данных

Использованы почасовые данные электропотребления горного предприятия Якутии за период с 01.01.2010 по 31.12.2013. В данных есть пропуски, поэтому общая выборка составляет 5366 суток, в среднем по 1 342 суток (32 208 часов) на одно предприятие. Отдельно выполнен анализ для обогатительной фабрики, нерюнгринского угольного разреза и предприятия в целом, включающего в себя кроме фабрики и разрезов дополнительных потребителей.

На рис. 3–5 приведены гистограммы распределения почасовых электропотреблений в кВт·ч. Примеры дневных графиков электропотребления за два следующих друг за другом дня показаны на рис. 6–8. Видно, что предприятия существенно статистически отличаются по своим процессам электропотребления.

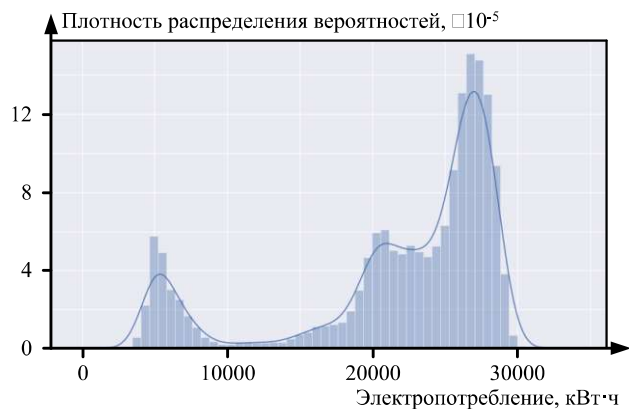


Рис. 3. Распределение электропотребления фабрик

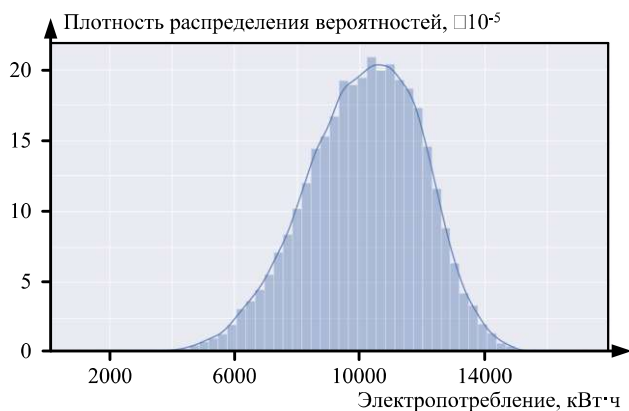


Рис. 4. Распределение электропотребления разреза



Рис. 5. Распределение электропотребления предприятия в целом

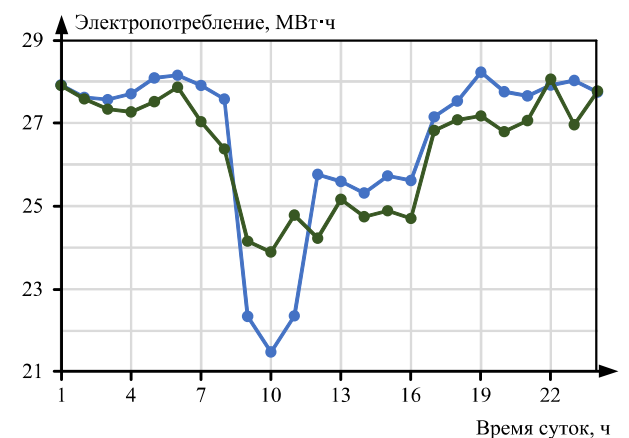


Рис. 6. Примеры дневных графиков нагрузки фабрики

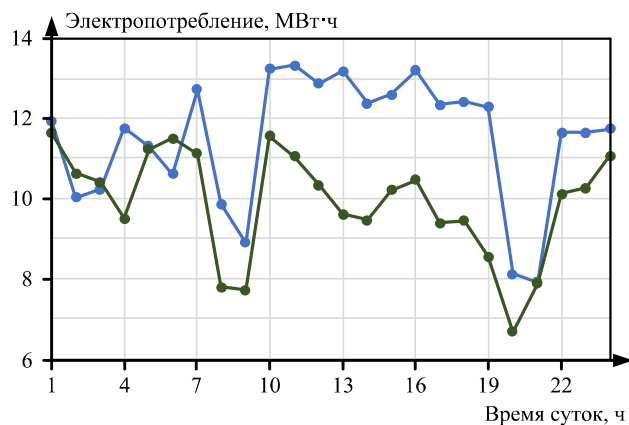


Рис. 7. Примеры дневных графиков нагрузки разреза



Рис. 8. Примеры дневных графиков нагрузки предприятия в целом

Отбор признаков

Важность отбора признаков для ансамблевых моделей в задачах электроэнергетики показана в работах [34-36]. На рис. 9 показано изменение коэффициента корреляции Спирмена между электропотреблением в заданный час и в предыдущие часы. Из рис. 9 видно, что наибольшая корреляция наблюдается между электропотреблением часа и трех предыдущих часов, но их использование невозможно, так как прогноз делается на сутки вперед.

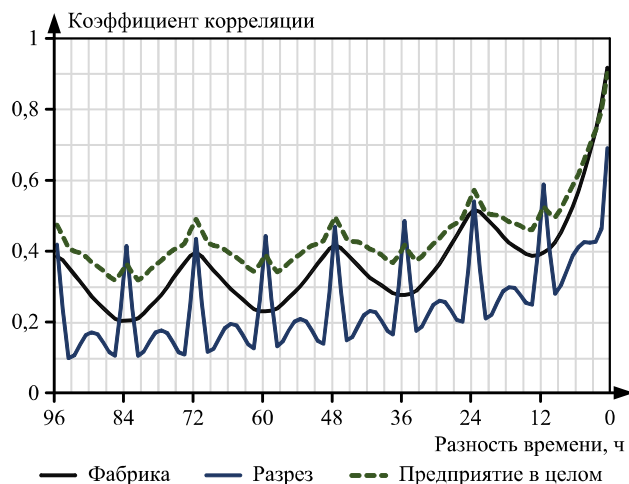


Рис. 9. Корреляции между электропотреблением в данный час и в предыдущие часы

Для всех предприятий выделяются часы, номера которых совпадают с рассматриваемым часом (часы, отстоящие от рассматриваемого на кратное число суток, 24, 48, 72, 96 ч). Для предприятия П₂ характерно, что значения электропотребления в часы, отстоящие друг от друга на $k + 0,5$ суток ($k = 0, 1, \dots$) тоже имеют повышенную корреляцию между собой. Можно выделить наиболее значимые для прогнозирования часы прошлых суток. Кроме того, в качестве значимых признаков выбраны номер часа суток, номер дня недели, число (номер дня месяца), месяц и год. Коэффициенты корреляции приведены в табл. 1. В табл. 1 столбцы «час_48» и «час_72» означают электропотребление в те же часы, что и прогнозный, но за 48 и 72 ч от него в прошлом.

Результаты применения моделей машинного обучения

Используемые в работе гиперпараметры ансамблевых моделей машинного обучения приведены в табл. 2. Подбор параметров выполнен с помощью метода случайного поиска (Random Search [37]). Кроме того, для сравнения результатов к используемым методам был добавлен классический вариант ИНС, многослойный перцептрон (MLP).

Выборка была случайным образом поделена на обучающую и тестовую в отношении 80 на 20. Число суток m из выражения (2) было экспериментально подобрано равным четырем. Значения гипер-параметров моделей приведены в табл. 2, а их результаты в табл. 3.

Усредненные по объектам ошибки прогнозирования показаны на рис. 10.

Таблица 1

Коэффициенты корреляции по критерию Спирмена

Объект	Час суток	День недели	Число	Месяц	Год	Час_48	Час_72
Фабрика	-0,07	0,02	0,04	0,22	0,31	0,52	0,42
Разрез	-0,09	-0,08	0	0,15	0,13	0,54	0,47
Все предприятия	-0,09	-0,01	0,06	0,23	0,39	0,5	0,57

Таблица 2

Гиперпараметры моделей

Модель	Гиперпараметры
Random Forest	Максимальная глубина дерева = 20, размер ансамбля = 120
AbaBoost	Максимальная глубина дерева = 9, размер ансамбля = 100
XGBoost	Максимальная глубина дерева = 8, размер ансамбля = 100
MLP	Число скрытых слоев = 2, нейронов в скрытых слоях = 200, 100 функции активации в скрытых слоях ReLU, метод обучения Adam

Таблица 3

Результаты применения моделей, средняя ошибка почасового прогноза на сутки вперед, MAPE, %

Модель	Фабрика, обуч.	Фабрика, тест	Разрез, обуч.	Разрез, тест	Предприятия в целом, обуч.	Предприятия в целом, тест
RF	8,6	20,2	4,5	10,6	5,0	11,6
AB	23,8	28,3	9,0	10,9	12,4	14,8
XGB	18,2	23,8	8,3	10,3	9,4	12,4
MLP	26,5	28,9	12,2	12,2	16,7	17,1
RNN	26,4	25,6	10,6	10,9	15,2	14,4

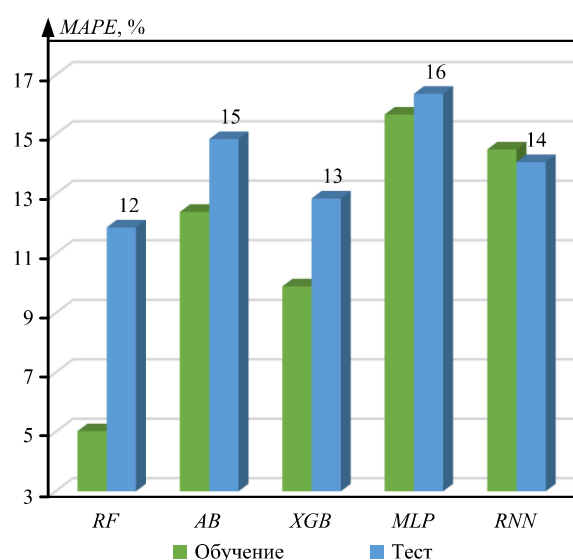


Рис. 10. Усредненная по всем объектам ошибка почасового прогноза на сутки вперед

Полученные результаты показывают следующее:

- обработка всего временного ряда электропотребления за прошлые m суток не дает существенного прироста точности по сравнению с использованием электропотребления только в отобранные наиболее значимые часы и может давать даже менее точный прогноз, так как разница ошибок RNN и Random Forest на тестовых выборках составила от 0,3 до 5,4 процентных пункта или от 3 до 24 %;

- рекуррентные ИНС способны самостоятельно обучаться выделению нужных признаков из всего временного ряда, в то время как для применения ансамблевых методов необходима внешняя процедура предварительного анализа и выделения признаков;

- за счет использования методик борьбы с переобучением для рекуррентных ИНС их точность на обучающей и тестовой выборках оказалась близкой, по этому показателю рекуррентные ИНС оказались существенно лучше ансамблевых методов;

- подтверждено, что для прогнозирования электропотребления рекуррентные ИНС более эффективны, чем классические многослойные полносвязные ИНС, так как разница ошибок RNN и MLP на тестовых выборках составила от 1,3 до 3,3 процентных пункта или от 11 до 30 %;

- в случае корректного применения методов машинного обучения и выбора класса моделей точность прогноза определяется скорее особенностями предприятия, чем выбором конкретной модели из класса;

- для достижения наилучшей точности прогноза необходимо применять различные ансамблевые методы и рекуррентные ИНС, поскольку заранее неизвестно, какой именно метод окажется лучше в конкретной задаче (согласно теореме No Free Lunch [38]), в данной задаче наилучший результат получен с помощью градиентного бустинга над деревьями решений.

На рис. 11 показаны прогнозные и фактические фрагменты графиков электропотребления в МВт·ч, полученные с помощью градиентного бустинга для наиболее сложного для прогнозирования объекта – обогатительной фабрики.

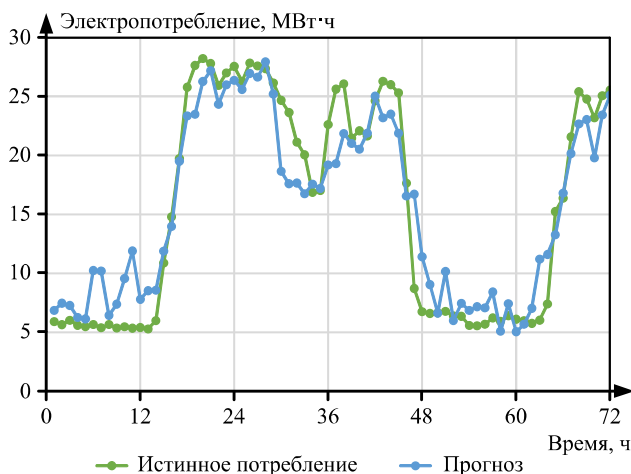


Рис. 11. Сопоставление прогнозов и истинных графиков электропотребления для фабрики

ЗАКЛЮЧЕНИЕ

Выполнено исследование ансамблевых и нейросетевых моделей машинного обучения в задаче краткосрочного прогнозирования почасовых графиков электропотребления горных предприятий. Использованы данные якутского горного предприятия (фабрики, разреза и предприятия в целом) за четырехлетний период. Показано, что в случае применения нейросетевых моделей более эффективным является рекуррентная архитектура, позволяющая обрабатывать данные об электропотреблении как временной ряд, учитывая динамику изменения электрической нагрузки.

В то же время за счет проведения предварительного анализа и отбора признаков возможно выбрать из временного наиболее значимые параметры и затем применить к ним ансамблевые модели машинного обучения, такие как случайный лес и адаптивный или градиентный бустинг над деревьями решений. Такие модели не способны учитывать динамику изменения нагрузки, но, как показало исследование, для прогнозирования электропотребления горных предприятий этого может и не требоваться.

Использование отбора признаков вместе с ансамблевыми методами дает близкую точность к рекуррентным нейронным сетям и может даже превосходить их. При этом такие модели обучаются на 1–2 порядка быстрее, но более склонны к переобучению.

Для рассмотренных потребителей средняя погрешность прогноза на сутки вперед составила от 10 до 20% и зависит в первую очередь от степени стохастичности технологических и геологических факторов горного предприятия.

Исследование выполнено при финансовой поддержке РФФИ, НТУ «Сирius», ОАО «РЖД» и Общественного Фонда «Талант и успех» в рамках научного проекта № 20-38-51007.

СПИСОК ЛИТЕРАТУРЫ

1. Data-driven baseline estimation of residential buildings for demand response / S. Park, S. Ruy, Y. Choi, J. Kim, H. Kim // *Energies*. 2015. Vol. 8(9). Pp. 10239–10259. doi: 10.3390/en80910239
2. Hahn H., Meyer-Nieberg S., Pickl S. Electric load forecasting methods: Tools for decision making // *European Journal of Operational Research*. 2009. Vol. 199. No. 3. Pp. 902–907.

doi: 10.1016/j.ejor.2009.01.062

3. A review on time series forecasting techniques for building energy consumption / C. Deb, F. Zhang, J. Yang, S. Lee, K. Shah // *Renewable and Sustainable Energy Reviews*. 2017. Vol. 74. Pp. 902–924.
4. Taylor J.W. Short-term electricity demand forecasting using double seasonal exponential smoothing // *The Journal of the Operational Research Society*. 2003. Vol. 54. No. 8. Pp. 799–805.
5. Short-Term Forecast of Electricity Load for LLC “Omsk Energy Retail Company” Using Neural Network / V. Potapov, R. Khamitov, V. Makarov, A. Gritsay, I. Chervenчук, D. Tyunkov // *Dynamics of Systems, Mechanisms and Machines (Dynamics)*. IEEE, 2018. Pp. 1–5. doi: 10.1109/Dynamics.2018.8601430
6. Cho M.Y., Hwang J.C., Chen C.S. Customer short term load forecasting by using ARIMA transfer function model // *International Conference on Energy Management and Power Delivery*. IEEE, 1995. Vol. 1. Pp. 317–322. doi: 10.1109/EMPD.1995.500746
7. Hagan M.T., Behr S.M. The time series approach to short term load forecasting. *IEEE Transactions on Power Systems*. 1987. Vol. 2(3). Pp. 785–791. doi: 10.1109/TPWRS.1987.4335210
8. Разработка моделей прогнозирования электропотребления на основе временных рядов в изолированных энергосистемах / Д.Х. Худжасидов, А.Г. Русина, П.В. Матренин, С.А. Дмитриев, М.Х. Сафаралиев // *Электротехнические системы и комплексы*. 2020. № 3(48). С. 23–27. doi: 10.18503/2311-8318-2020-3(48)-23-27
9. Short-term load forecasting for the holidays using fuzzy linear regression method / K.B. Song, Y.S. Baek, D.H. Hong, G. Jang // *IEEE Transactions on Power Systems*. 2005. Vol. 20. No. 1. Pp. 96–101. doi: 10.1109/PES.2005.1489152
10. Antonenkov D.V., Solovlev D.B. Mathematic simulation of mining company’s power demand forecast (by example of “Neryungri” coal strip mine) // *IOP Conference Series: Earth and Environmental Science*. 2017. Vol. 87. Art. no. 032003. doi: 10.1088/1755-1315/87/3/032003
11. Hong W.C. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm // *Energy*. 2011. Vol. 36. No. 9. Pp. 5568–5578. doi: 10.1016/j.energy.2011.07.015
12. Hybrid forecasting approach based on GRNN neural network and SVR machine for electricity demand forecasting / W. Li, X. Yang, H. Li, L. Su // *Energies*. 2017. Vol. 10. No. 1. Art. no. 44. doi: 10.3390/en10010044
13. Huo J., Shi T., Chang J. Comparison of random forest and SVM for electrical short-term load forecast with different data sources // *7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2016. Pp. 1077–1080. doi: 10.1109/ICSESS.2016.7883252
14. A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts / R. Zhang, Y. Xu, Z.Y. Dong, W. Kong, K.P. Wong // *Power and Energy Society General Meeting*. IEEE, 2016. Pp. 1–5. doi: 10.1109/PESGM.2016.7741097
15. Mori H., Kosemura N. Optimal regression tree based rule discovery for short-term load forecasting // *Power Engineering Society Winter Meeting*. IEEE, 2001. Vol. 2. Pp. 421–426. doi: 10.1109/PESW.2001.916878
16. Ensemble deep learning for regression and time series forecasting / X. Qiu, L. Zhang, Y. Ren, P. Suganthan, G. Amaratunga // *Symposium on Computational Intelligence in Ensemble Learning*. IEEE, 2014. Pp. 21–26. doi: 10.1109/CIEL.2014.7015739
17. Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting / Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, Z. Zhu // *IEEE Access*. 2019. Vol. 7. Pp. 88058–88071. doi: 10.1109/ACCESS.2019.2926137

18. A short-term load forecasting method based on GRU-CNN hybrid neural network model / L. Wu, C. Kong, X. Hao, W. Chen // *Mathematical Problems in Engineering*. 2020. Art. no. 1428104. doi: 10.1155/2020/1428104
19. Ryu S., Noh J., Kim H. Deep neural network based demand side short term load forecasting // *Energies*. 2016. Vol. 10. No. 1. Pp. 1–20. doi: 10.3390/en10010003
20. Chen K. Wang Q., He Z., Hu J., He J. Short-Term Load Forecasting with Deep Residual Networks. *IEEE Trans. Smart Grid*. 2019, vol. 10, no. 4, pp. 3943–3952. doi: 10.1109/TSG.2018.2844307
21. Cai M., Pipattanasomporn M., Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques // *Applied Energy*. 2019. Vol. 236. Pp. 1078–1088. doi: 10.1016/j.apenergy.2018.12.042
22. Short-term residential load forecasting based on LSTM recurrent neural network / W. Kong, Z. Dong, Y. Jia, D. Hill, Y. Xu, Y. Zhang // *IEEE Transactions on Smart Grid*. 2019. Vol. 10. No. 1. Pp. 841–851. doi: 10.1109/TSG.2017.2753802
23. Improving accuracy and generalization performance of small-size recurrent neural networks applied to short-term load forecasting / P.V. Matrenin, V.Z. Manusov, A.I. Khal-yasmaa, D.V. Antonenkov, S.A. Eroshenko, D. Butusov // *Mathematics*. 2020. Vol. 8. No. 12. Art. no. 2169. doi: 10.3390/math8122169
24. Hochreiter S., Schmidhuber J. Long short-term memory // *Neural Computation*. 1997. Vol. 9. No. 8. Pp. 1735–1780. doi: 10.1162/neco.1997.9.8.1735
25. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation / K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio // Available online: <https://arxiv.org/abs/1406.1078>
26. Le Cun Y., Bengio Y., Hinton G. // *Nature*. 2015. Vol. 521. Pp. 436–444. doi: 10.1038/nature14539
27. Dahl G.E., Sainath T.N., Hinton G.E. Improving deep neural networks for LVCSR using rectified linear units and dropout. *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013. Pp. 8609–8613. doi: 10.1109/ICASSP.2013.6639346
28. Ng A.Y. Feature selection. L1 vs. L2 regularization. and rotational invariance // *21st International Conference on Machine Learning*. Banff, 2004. Pp. 1–8. doi: 10.1145/1015330.1015435
29. Kingma D.P., Ba J.L. Adam: A method for stochastic optimization. [Web source]. Available online: <https://arxiv.org/abs/1412.6980>
30. Breiman L. Random Forests // *Machine Learning*. 2001. Vol. 4. Pp. 5–32. doi: 10.1023/A:1010933404324
31. Machine learning in Python. [Web source]. Available online: <https://scikit-learn.org>
32. Drucker H. Improving Regressors using Boosting Techniques. [Web source]. Available online: <http://www.researchgate.net/publication/2424244>
33. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. [Web source]. Available online: <https://arxiv.org/abs/1603.02754>
34. Industry experience of developing day-ahead photovoltaic plant forecasting system based on machine learning / Khal-yasmaa A.I., Eroshenko S.A., Tashchilin V.A., Ramachandran H., Chakravarthi T.P., Butusov D.N. // *Remote Sensing*. 2020. Vol. 12(20). Art. no. 3420. doi: 10.3390/rs12203420
35. Khalyasmaa A.I., Senyuk M.D., Eroshenko S.A. Analysis of the state of high-voltage current transformers based on gradient boosting on decision trees // *IEEE Transactions on Power Delivery*. 2020. Vol. 36. No. 4. Pp. 2154–2163. doi: 10.1109/TPWRD.2020.3021702
36. Разработка моделей среднесрочного прогнозирования электропотребления в изолированно работающих энергосистемах на основе ансамблевых методов машинного обучения / С.М. Асанова, Д.С. Ахьеев, С.А. Дмитриев, П.В. Матренин, М.Х. Сафаралиев // *Известия НТЦ Единой энергетической системы*. 2021. № 1(84). С. 32–39.
37. Bergstra J., Bengio Y. Random search for hyper-parameter optimization // *Journal of Machine Learning Research*. 2012. Vol. 13. Pp. 281–305.
38. Wolpert D.H., Macready W.G. No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation*. 1997. Vol. 1. No. 1. Pp. 67–82. doi: 10.1109/4235.585893

Поступила в редакцию 10 июня 2021 г.

INFORMATION IN ENGLISH

ENSEMBLE AND NEURAL NETWORK MACHINE LEARNING MODELS FOR SHORT-TERM LOAD FORECASTING OF OPEN CAST MINING COMPANIES

Dmitry V. Antonenkov

Ph.D. (Engineering), Associate Professor, Department of Industrial Power Supply Systems, Novosibirsk State Technical University, Novosibirsk, Russia. E-mail: antonenkov@corp.nstu.ru

Pavel V. Matrenin

Ph.D. (Engineering), Associate Professor, Department of Industrial Power Supply Systems, Novosibirsk State Technical University, Novosibirsk, Russia. E-mail: matrenin.2012@corp.nstu.ru. ORCID: 0000-0001-5704-0976

The article deals with the problem of forecasting the power consumption at mining enterprises, which is characterized by a high nonstationarity and stochasticity levels. The data collected by the authors for four years on the mining enterprise of Yakutia, working in the field of coal mining and processing, are used. At the same time, a separate analysis was carried out for various objects of the enterprise: a coal mine and processing plants, which have fundamentally different technological processes, and, consequently, power consumption schedules. A study of two classes of machine learning methods was carried out: processing of retrospective data on the power consumption at an enterprise as a time series using recurrent neural networks; and extraction of

the most significant features in order to apply to them ensemble models based on decision trees: a random forest, adaptive boosting and extreme gradient boosting. Since tuning hyper-parameters is very important for the specified machine learning models, for the correct comparison of the results, the procedure for optimizing the hyper-parameters of all models was carried out. The computational experiments have shown that recurrent multilayer neural networks are able to use time series for forecasting without preliminary processing, learning to recognize significant signs from the dynamics of changes in the electrical consumption schedule. To apply ensembles of regression decision trees, preliminary data analysis is required to extract the most

significant features from a time series. Using the example of the considered enterprise, it is shown that the use of such an approach when working with ensemble models gives an accuracy close to that of recurrent neural networks. In this case, ensemble models are trained 1–2 orders of magnitude faster, and the disadvantage is a great tendency to overfitting.

Keywords: opencast mining, power consumption forecasting, machine learning, recurrent neural network, feature selection, ensemble models.

REFERENCES

1. Park S., Ruy S., Choi Y., Kim J., Kim H. Data-driven baseline estimation of residential buildings for demand response. *Energies*, 2015, vol. 8, pp. 10239–10259. doi: 10.3390/en80910239
2. Hahn H., Meyer-Nieberg S., Pickl S. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, 2009, vol. 199, no. 3, pp. 902–907. doi: 10.1016/j.ejor.2009.01.062
3. Deb C., Zhang F., Yang J., Lee S., Shah K. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 2017, vol. 74 (C), pp. 902–924.
4. Taylor J.W. Short-term electricity demand forecasting using double seasonal exponential smoothing. *The Journal of the Operational Research Society*, 2003, vol. 54, no. 8, pp. 799–805.
5. Potapov V., Khamitov R., Makarov V., Gritsay A., Cherven-chuk I., Tyunkov D. Short-Term Forecast of Electricity Load for LLC "Omsk Energy Retail Company" Using Neural Network. 2018 Dynamics of Systems, Mechanisms and Machines (Dynamics). Omsk, 2018, pp. 1–5. 10.1109/Dynamics.2018.8601430
6. Cho M.Y., Hwang J.C., Chen C.S. Customer short term load forecasting by using ARIMA transfer function model. *International Conference on Energy Management and Power Delivery*. Singapore, 1995, vol. 1, pp. 317–322. doi: 10.1109/EMPD.1995.500746
7. Hagan M.T., Behr S.M. The time series approach to short term load forecasting. *IEEE Trans. Power Syst*, 1987, vol. 2, pp. 785–791. doi: 10.1109/TPWRS.1987.4335210
8. Khujaaidov J.Kh., Rusina A.G., Matrenin P.V., Dmitriev S.A., Safaraliev M.Kh. Forecasting model development for electricity consumption based on time. *Elektrotehnicheskie sistemy i komplekсы* [Electrotechnical Systems and Complexes], 2020, no. 3 (48), pp. 23–27. (In Russian). doi: 10.18503/2311-8318-2020-3(48)-23-27
9. Song K.B., Baek Y.S., Hong D.H., Jang G. Short-term load forecasting for the holidays using fuzzy linear regression method. *IEEE Trans. Power Syst*, 2005, vol. 20, no. 1, pp. 96–101. doi: 10.1109/PES.2005.1489152
10. Antonenkov D.V., Solovov D.B. Mathematic simulation of mining company's power demand forecast (by example of "Neryungri" coal strip mine). *IOP Conf. Series: Earth and Environmental Science*, 2017, vol. 87, Art. no. 032003. doi: 10.1088/1755-1315/87/3/032003
11. Hong W.C. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy*, 2011, vol. 36, no. 9, pp. 5568–5578. doi: 10.1016/j.energy.2011.07.015
12. Li W., Yang X., Li H., Su L. Hybrid forecasting approach based on GRNN neural network and SVR machine for electricity demand forecasting. *Energies*, 2017, vol. 10, no. 1, Art. id 44. doi: 10.3390/en10010044
13. Huo J., Shi T., Chang J. Comparison of Random Forest and SVM for electrical short-term load forecast with different data sources. *IEEE International Conference on Software Engineering and Service Science*, Beijing. IEEE, 2016, pp. 1077–1080. doi: 10.1109/PESGM.2016.7741097
14. Zhang R., Xu Y., Dong Z.Y., Kong W., Wong, K.P. A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts. 2016 IEEE Power and Energy Society General Meeting (PESGM). IEEE, 2016, pp. 1–5. doi: 10.1109/PESGM.2016.7741097
15. Mori H., Kosemura N. Optimal regression tree based rule discovery for short-term load forecasting. 2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194). IEEE, 2001, vol. 2, pp. 421–426. doi: 10.1109/PESW.2001.916878
16. Qiu X., Zhang L., Ren Y., Suganthan P., Amaratunga G. Ensemble deep learning for regression and time series forecasting. 2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL). IEEE, 2014, pp. 21–26. doi: 10.1109/CIEL.2014.7015739
17. Deng Z., Wang B., Xu Y., Xu T., Liu C., Zhu Z. Multi-Scale Convolutional Neural Network with Time-Cognition for Multi-Step Short-Term Load Forecasting. *IEEE Access*, 2019, vol. 7, pp. 88058–88071. doi: 10.1109/ACCESS.2019.2926137
18. Wu, L.; Kong, C.; Hao, X.; Chen, W. A Short-Term Load Forecasting Method Based on GRU-CNN Hybrid Neural Network Model. *Mathematical Problems in Engineering*, 2020, Art. no. 1428104. doi: 10.1155/2020/1428104
19. Ryu S., Noh J., Kim H. Deep neural network based demand side short term load forecasting. *Energies*, 2016, vol. 10, no. 1, Art. id 3, pp. 1–20. doi: 10.3390/en10010003
20. Chen K., Wang Q., He Z., Hu J., He J. Short-Term Load Forecasting with Deep Residual Networks. *IEEE Trans. Smart Grid*, 2019, vol. 10, no. 4, pp. 3943–3952. doi: 10.1109/TSG.2018.2844307
21. Cai M., Pipattanasomporn M., Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Applied Energy*, 2019, vol. 236, pp. 1078–1088. doi: 10.1016/j.apenergy.2018.12.042
22. Kong W., Dong Z., Jia Y., Hill D., Xu Y., Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans. on Smart Grid*, 2019, vol. 10, no. 1, pp. 841–851. doi: 10.1109/TSG.2017.2753802
23. Matrenin P.V., Manusov V.Z., Khalyasmaa A.I., Antonenkov D.V., Eroshenko S.A., Butusov D. Improving accuracy and generalization performance of small-size recurrent neural networks applied to short-term load forecasting. *Mathematics*, 2020, vol. 8, no. 12, Art. id 2169. doi: 10.3390/math8122169
24. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, vol. 9, no. 8, pp. 1735–1780. doi: 10.1162/neco.1997.9.8.1735
25. Cho K., van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Available at: <https://arxiv.org/abs/1406.1078>
26. LeCun Y., Bengio Y., Hinton G. *Nature*, 2015, vol. 521, pp. 436–444. doi: 10.1038/nature14539
27. Dahl G.E., Sainath T.N., Hinton G.E. Improving deep neural networks for LVCSR using rectified linear units and dropout. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013, pp. 8609–8613. doi: 10.1109/ICASSP.2013.6639346
28. Ng A.Y. Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the twenty-first international conference on Machine learning (ICML '04)*. Association for Computing Machinery, 2004, pp. 1–8. doi: 10.1145/1015330.1015435
29. Kingma D.P., Ba J.L. Adam: A method for stochastic optimization. Available at: <https://arxiv.org/abs/1412.6980>
30. Breiman L. Random Forests. *Machine Learning*. 2001. Vol. 4. Pp. 5–32. doi: 10.1023/A:1010933404324.
31. Machine learning in Python. Available at: <https://scikit-learn.org>
32. Drucker H. Improving Regressors using Boosting Techniques.

- Available at: <https://www.researchgate.net/publication/2424244>
33. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. Available at: <https://arxiv.org/abs/1603.02754>
 34. Khalyasmaa A.I., Eroshenko S.A., Tashchilin V.A., Ramachandran H. Chakravarthi T.P., Butusov D.N. Industry Experience of Developing Day-Ahead Photovoltaic Plant Forecasting System Based on Machine Learning. Remote Sensing, 2020, vol. 12 (20), Art. id 3420. doi: 10.3390/rs12203420
 35. Khalyasmaa A.I., Senyuk M.D., Eroshenko S.A. Analysis of the state of high-voltage current transformers based on gradient boosting on decision trees. IEEE Transactions on Power Delivery, 2020, vol. 36, no. 4, pp. 2154-2163. doi: 10.1109/TPWRD.2020.3021702
 36. Asanova S.M., Ahyeev D.S. Dmitriev S.A., Matrenin P.V., Safaraliev M.Kh. Development of models for power consumption medium-term forecasting in isolated power systems based on ensemble methods of machine learning. *Izvestiya NTTS Edinoy energeticheskoy sistemy* [STC of Unified Power System Proceedings], 2021, no. 1(84), pp. 32-39. (In Russian)
 37. Bergstra J., Bengio Y. Random search for hyper-parameter optimization. Journal of Machine Learning Research, 2012, vol. 13, pp. 281-305.
 38. Wolpert D.H., Macready W.G. No Free Lunch Theorems for Optimization. IEEE Transactions on Evolutionary Computation, 1997, vol. 1, no. 1, pp. 67-82. doi: 10.1109/4235.585893

Антоненков Д.В., Матренин П.В. Исследование ансамблевых и нейросетевых методов машинного обучения в задаче краткосрочного прогнозирования электропотребления горных предприятий // Электротехнические системы и комплексы. 2021. № 3(52). С. 57-65. [https://doi.org/10.18503/2311-8318-2021-3\(52\)-57-65](https://doi.org/10.18503/2311-8318-2021-3(52)-57-65)

Antononkov D.V., Matrenin P.V. Ensemble and Neural Network Machine Learning Models for Short-Term Load Forecasting of Open Cast Mining Companies. *Elektrotekhnicheskie sistemy i komplekсы* [Electrotechnical Systems and Complexes], 2021, no. 3(52), pp. 57-65. (In Russian). [https://doi.org/10.18503/2311-8318-2021-3\(52\)-57-65](https://doi.org/10.18503/2311-8318-2021-3(52)-57-65)
