

ВЫБОР ОПТИМАЛЬНОЙ АРХИТЕКТУРЫ И КОНФИГУРАЦИИ НЕЙРОСЕТЕВОЙ МОДЕЛИ ПРИ КРАТКОСРОЧНОМ ПРОГНОЗИРОВАНИИ ЭЛЕКТРОПОТРЕБЛЕНИЯ ГАРАНТИРУЮЩИХ ПОСТАВЩИКОВ ЭЛЕКТРОЭНЕРГИИ

Серебряков Николай Александрович, na_serebryakov@altke.ru

Аннотация:

Для эффективного управления режимом работы энергосистемы, необходима прогнозная информация о величине почасовой электрической нагрузки всех потребителей. Ошибки прогнозирования, как правило, приводят к ухудшению технологических и экономических показателей функционирования энергосистемы, за счет необоснованного изменения режима работы генерирующего оборудования, а также выборы неоптимальной схемы электрических сетей. Данная статья посвящена вопросам повышения точности краткосрочного прогнозирования почасового электропотребления (short-term load forecasting – STLF) группы точек поставки гарантирующего поставщика (ГТП ГП) с помощью инструментов искусственных нейронных сетей. Важнейшим условием достижения высокой точности STLF является выбор оптимальной архитектуры и конфигурации прогнозной нейросетевой модели. Произведен анализ эффективности применения стохастического градиентного спуска, а также его современных модификаций, для оптимизации функции энергии ошибки нейронной сети. В ходе экспериментов с обучающей и тестовой выборкой данных, выяснено, что наибольшую точность STLF демонстрируют нейросетевые модели, оптимизируемые с помощью метода адаптивной инерции (adaptive momentum estimation – ADAM). Выяснено, что современные инструменты глубокого машинного обучения, такие как алгоритм HyperBand, позволяют автоматизировать процесс оптимизации гиперпараметров нейросетевой модели. С помощью алгоритма HyperBand подобраны оптимальные значения гиперпараметров для многослойного персептрона, одномерной и двухмерной сверточных нейросетей, рекуррентной нейросети, а также ансамбля, состоящего из вышеперечисленных нейросетей. В ходе сравнительного анализа точности краткосрочного прогнозирования почасового электропотребления ГТП ГП, полученного с помощью различных нейросетевых алгоритмов, выяснено, что ансамбль нейросетей показывает наибольшую точность прогнозирования на обучающей и тестовой выборках данных. Применение ансамбля нейросетей при прогнозировании почасового электропотребления ГТП ГП позволило снизить среднюю абсолютную процентную ошибку прогнозирования на 0,22% на месячной интервале и на 0,12 % на годовом интервале, по сравнению с прогнозом, полученным с помощью метода экспертных оценок.

Ключевые слова: группа точек поставки, краткосрочное прогнозирование электропотребления, искусственная нейронная сеть, стохастическая оптимизация, алгоритм обучения, гиперпараметр, ошибка прогнозирования, кривые обучения.

Необходимым условием надежного функционирования энергосистемы любого уровня является поддержание постоянного баланса между выработкой и потреблением электрической энергии. Так как, пуск большей части генерирующего оборудования в энергосистеме занимает более шести часов, то для эффективного управления режимом работы энергосистемы необходим краткосрочный почасовой прогноз электропотребления всех потребителей (short-term

load forecasting – STLF). В условиях функционирования рынка электроэнергии точность прогнозов потребления существенно влияет на технологические и экономические показатели энергосистемы [1]. Ошибки прогноза потребления электроэнергии ведут к необоснованным пускам и остановам генерирующего оборудования, а также выбору неоптимальной схемы электрических сетей.

Почасовое электропотребление группы точек поставки (ГТП) гарантирующего поставщика электроэнергии (ГП) зависит как от стандартных временных и метеорологических факторов, так и от факторов:

- плановых и аварийных отключений питающих электрических сетей 6-110 кВ;
- графика функционирования потребителей электроэнергии с максимальной мощностью свыше 670 кВт;
- наличие или отсутствие центрального отопления и водоснабжения в населенных пунктах, питающихся от данной группы точек поставки.

В связи с учетом дополнительных факторов при прогнозе электропотребления ГТП ГП, остаются открытыми вопросы выбора оптимальной архитектуры и конфигурации прогнозного алгоритма. В условиях неопределенности традиционные методы математической статистики или имитационного моделирования не позволяют строить адекватные модели объектов [2]. В настоящее время, активно развиваются технологии искусственного интеллекта и машинного обучения. Рассмотренные информационные технологии являются высокоэффективным инструментом для решения широкого перечня задач, которые относятся к слабоформализуемым или неформализуемым [3].

Статистические подходы к STLF обычно нуждаются в математической модели, которая представляет электрическую нагрузку как функцию различных факторов, таких как время, погода и класс потребителей [4]. В работе [5], авторами предложено представить почасовую электрическую нагрузку ГТП ГП в виде скалярного произведения столбца влияющих факторов и матрицы свободных коэффициентов. В этом случае, задача краткосрочного прогнозирования электропотребления сводится к поиску значений матрицы свободных коэффициентов с помощью инструментов глубокого машинного обучения.

Проектирование нейронной сети, предназначенной для краткосрочного прогнозирования электропотребления гарантирующего поставщика, включает в себя оптимизацию в различных ее проявлениях. Свободные параметры нейросети оптимизируются в процессе обучения на статистических данных. К данным параметрам относятся весовые коэффициенты взаимодействия между нейронами и пороги активации. Обучение современных глубоких искусственных нейронных сетей (ИНС), предназначенных для краткосрочного прогнозирования электропотребления, происходит на основании алгоритма обратного распространения ошибки. В процессе работы данного алгоритма происходит два прохода вычислений: прямой и обратный. Во время прямого прохода вычислений на вход нейросети подается вектор признаков. После прохождения сигнала через все слои нейросети формируется выходной сигнал, являющийся откликом нейронной сети на входные воздействие. При прямом проходе сигнала свободные параметры сети остаются неизменными.

Далее, во время обратного прохода, рассчитывается функция энергии ошибки по всем нейронам выходного слоя:

$$J(n) = \frac{1}{2C} \sum_{j \in C} e_j^2(n), \quad (1)$$

где $J(n)$ — функция стоимости на итерации обучающего цикла n ; C — набор нейронов, относящихся к выходному слою; e_j^2 — квадратичная ошибка нейрона j .

В данном случае имеется задача оптимизации, целевым параметром которой является функция стоимости. Наиболее успешным, на сегодняшний день, алгоритмом оптимизации сложных функций является алгоритм градиентного спуска (gradient descent algorithm). Если целевая функция дифференцируема по своим параметрам, то градиентный спуск является относительно эффективным методом оптимизации [6].

В случае использования алгоритма градиентного спуска для оптимизации функции стоимости $J(n)$, данная функция представляется в виде функциональной зависимости синоптических весов нейронной сети $J(w)$. Для визуализации процесса обучения строится поверхность ошибки функции $J(w)$ в пространстве свободных параметров нейронной сети. Основная идея алгоритма обратного распространения ошибки состоит в коррекции синоптического веса $w_{ij}(n)$ на величину $\Delta w_{ij}(n)$, которая пропорциональна градиенту функции стоимости $\frac{\partial J(n)}{\partial w_{ij}(n)}$:

$$\nabla J(n) = \frac{\partial J(n)}{\partial w_{ij}(n)} = \frac{\partial J(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ij}(n)} = -e_j(n) \phi'_j(v_j(n)) y_i(n), \quad (2)$$

где $\nabla J(n)$ — градиент функции потерь на шаге обучения n ; $w_{ij}(n)$ — вес синаптической связи между нейронами i и j ; $e_j(n)$ — сигнал ошибки нейрона j на шаге обучения n ; $y_j(n)$ — отклик нейрона j на шаге обучения n ; $\partial v_j(n)$ — индуцированное локальное поле нейрона j на шаге обучения n ; ϕ'_j — функция активации.

Коррекция, применяемая к синоптическому весу взаимодействия $w_{ij}(n)$ определяется согласно дельта-правилу:

$$\Delta w_{ij}(n) = -\eta \frac{\partial J(n)}{\partial w_{ij}(n)} = \eta e_j(n) \phi'_j(v_j(n)) y_i(n), \quad (3)$$

где $\Delta w_{ij}(n)$ — величина изменения веса синаптической связи между нейронами i и j на шаге обучения n ; η — коэффициент скорости обучения сети.

Как мы видим из выражения (3), целью обучения нейронной сети является итеративная коррекция весов синоптических связей ИНС, с целью нахождения глобального минимума функции стоимости:

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n). \quad (4)$$

Стохастический градиентный спуск работает без проблем при оптимизации выпуклых целевых функций. Однако, функция энергии ошибки многослойной нейронной сети в большинстве случаев будет являться невыпуклой функцией. То есть, данная функция имеет множество локальных минимумов, плато и седловых точек, в окрестности которых градиент принимает нулевое или положительные значения. Данное обстоятельство способствует замедлению сходимости алгоритма градиентного спуска в область глобального минимума поверхности ошибки. Для ускорения сходимости алгоритма обучения в выражение (3) добавляется член момента инерции:

$$\Delta w_{ij}(n) = \eta e_j(n) \phi'_j(v_j(n)) y_i(n) + \mu \Delta w_{ij}(n-1), \quad (5)$$

где μ — параметр момента инерции; $\Delta w_{ij}(n-1)$ — корректировка веса синоптической связи между нейронами i и j на итерации обучения $n-1$.

Накопление импульса, за счет применения параметра момента инерции, хоть и улучшает сходимость алгоритма обучения, но появляется дополнительный параметр нейросети, который невозможно определить в процессе обучения. Также, применение момента не гарантирует достижения рабочей точки алгоритма глобального минимума на поверхности ошибки.

За последние несколько лет зарубежные ученые разработали несколько алгоритмов, минимизирующих количество параметров, которые требуют настройки в процессе обучения нейросети. Современные инструменты машинного обучения позволяют реализовать алгоритм обратного распространения ошибки с адаптивным, в процессе обучения, параметром скорости обучения, такие как ADAM и RMSProp [6].

Алгоритм RMSProp это одна из модификаций алгоритма градиентного спуска (SGD), призванная улучшить сходимость SGD при невыпуклом характере функции потерь J . Идея алгоритма RMSProp заключается в коррекции параметра скорости обучения за счет накопления в знаменателе экспоненциально взвешенного скользящего среднего из квадратов градиентов функции потерь J (root mean square propagation — RMSprop):

$$E[g^2(n)] = \gamma E[g^2(n-1)] + (1-\gamma)g^2(n), \quad (6)$$

где $E[g^2(n)]$ — экспоненциально взвешенное скользящее среднее из квадратов градиентов функции потерь J на шаге обучающего цикла n ; γ — сглаживающая константа, значение которой по умолчанию равно 0,9; $g(n) = \nabla J(n)$ — градиент функции потерь на шаге обучающего цикла n .

Тогда дельта-правило коррекции свободных параметров нейронной сети при использовании алгоритма RMSProp принимает вид:

$$w_{ij}(n+1) = w_{ij}(n) + \frac{\eta}{\sqrt{E[g^2(n)] + \varepsilon}} g(n), \quad (7)$$

где ε — константа небольшого значения, необходимая для исключения деления на ноль на начальных шагах обучающего цикла.

Алгоритм ADAM (adaptive momentum estimation — оценка адаптивных моментов) представляет собой комбинацию алгоритмов RMSProp и импульсного метода. Во-первых, рассчитывается первый момент, который представляет собой экспоненциальное скользящее среднее значений градиента функции потерь J и его скорректированное значение:

$$\begin{cases} m(n) = \beta_1 m(n-1) + (1 - \beta_1) g(n), \\ \hat{m}(n) = \frac{m(n)}{1 - \beta_1^n}, \end{cases} \quad (8)$$

где $m(n)$ — экспоненциальное скользящее среднее значений градиента функции потерь J ; β_1 — сглаживающая константа, значение которой по умолчанию равно 0,9; $\hat{m}(n)$ — скорректированное значение экспоненциального скользящего среднего значений градиента функции потерь J .

Далее рассчитывается второй момент, который представляет из себя среднюю нецентрированную дисперсию функции потерь J и ее скорректированное значение:

$$\begin{cases} v(n) = \beta_2 v(n-1) + (1 - \beta_2) g^2(n), \\ \hat{v}(n) = \frac{v(n)}{1 - \beta_2^n}, \end{cases} \quad (9)$$

где $v(n)$ — нецентрированная дисперсия функции потерь J ; β_2 — сглаживающая константа, значение которой по умолчанию равно 0,999; $\hat{v}(n)$ — скорректированное значение нецентрированной дисперсии функции потерь J .

Корректировка моментов $m(n)$ и $v(n)$ производится для искусственного увеличения данных параметров на начальных итерациях обучающего цикла.

Дельта-правило коррекции свободных параметров нейронной сети при использовании алгоритма ADAM принимает вид:

$$w_{ij}(n+1) = w_{ij}(n) + \frac{\eta}{\sqrt{\hat{v}(n) + \varepsilon}} \hat{m}(n). \quad (10)$$

Использование адаптивных оптимизаторов функции энергии ошибки способствует улучшению сходимости алгоритма обучения к окрестности глобального минимума вне зависимости от «ландшафта» поверхности ошибки. В данном случае, рабочая точка алгоритма обучения более эффективно проходит локальные минимумы на поверхности ошибки. Алгоритмы ADAM и RMSProp хорошо зарекомендовали себя на практике и окончательный выбор происходит на основании экспериментов с обучающей и проверочной выборками.

На рисунке 1 представлены кривые изменения средней абсолютной процентной ошибки прогнозирования почасового электропотребления ГТП ГП (кривые обучения) при использовании различных оптимизаторов функции энергии ошибки.

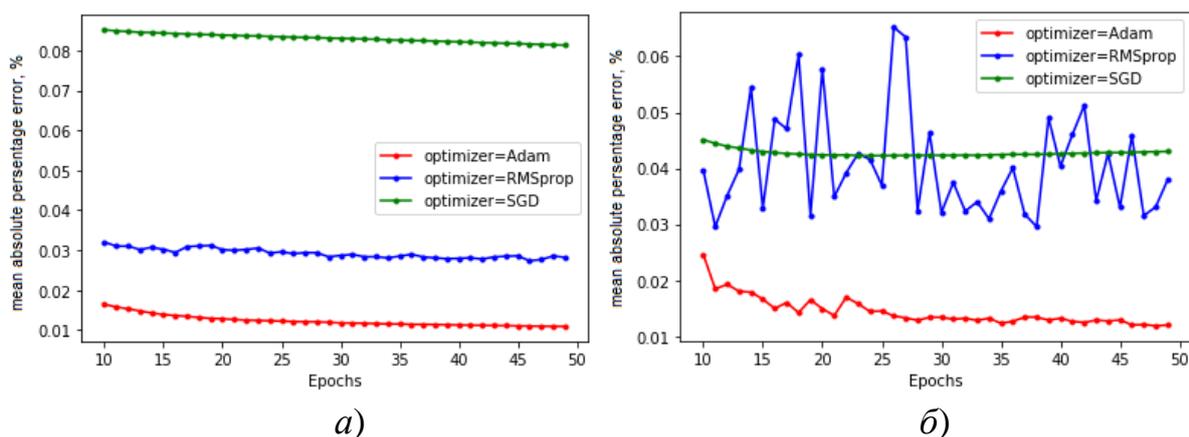


Рисунок 1 - Кривые изменения средней абсолютной процентной ошибки прогнозирования почасового электропотребления ГТП ГП при использовании различных оптимизаторов функции энергии ошибки на: а) обучающей выборке; б) тестовой выборке

Остальные параметры нейросети, которые называются гиперпараметрами, не могут быть оптимизированы в процессе обучения. К ним относятся:

- количество скрытых слоев;
- размер каждого слоя;
- вид функции активации.

До недавнего времени, единственным способом оптимизации гиперпараметров являлся простой перебор различных вариантов архитектур, типов нейронов, начальных условий, методов обучения и дальнейшим выбором наиболее удачного решения [7]. Однако, современные инструменты машинного обучения позволяют автоматизировать процесс подбора гиперпараметров. Одним из наиболее успешных алгоритмов оптимизации гиперпараметров нейронных сетей является алгоритм Hyperband [8]. Суть работы данного алгоритма заключается в следующем. В начале работы данного алгоритма задается набор гиперпараметров, оптимальное значение которых необходимо найти, диапазон изменения каждого гиперпараметра и дискретность шага поиска:

- искомый гиперпараметр: размер первого скрытого слоя;
- диапазон изменения гиперпараметра: [20;1000];
- шаг поиска: 10.

Далее, моделируются нейросетевые модели всевозможной конфигурации гиперпараметров из заданного диапазона. Затем, в процессе обучения, после некоторого количества эпох, отбрасывается половина конфигураций нейросетей, показавших худшую точность прогнозирования. Процесс продолжается, пока не останутся 2-3 наилучшие конфигурации нейросети. На рисунке 2 представлены кривые обучения нейросетевых моделей при работе алгоритма Hyperband.

С помощью алгоритма Hyperband определена оптимальная конфигурация для:

- многослойного перцептрона (multilayer perceptron – MLP);
- одномерной сверточной нейросети (convolution neural network CNN);
- трехмерной сверточной нейросети;

- рекуррентной нейросети типа долго-краткосрочной памяти (long-short term memory – LSTM).

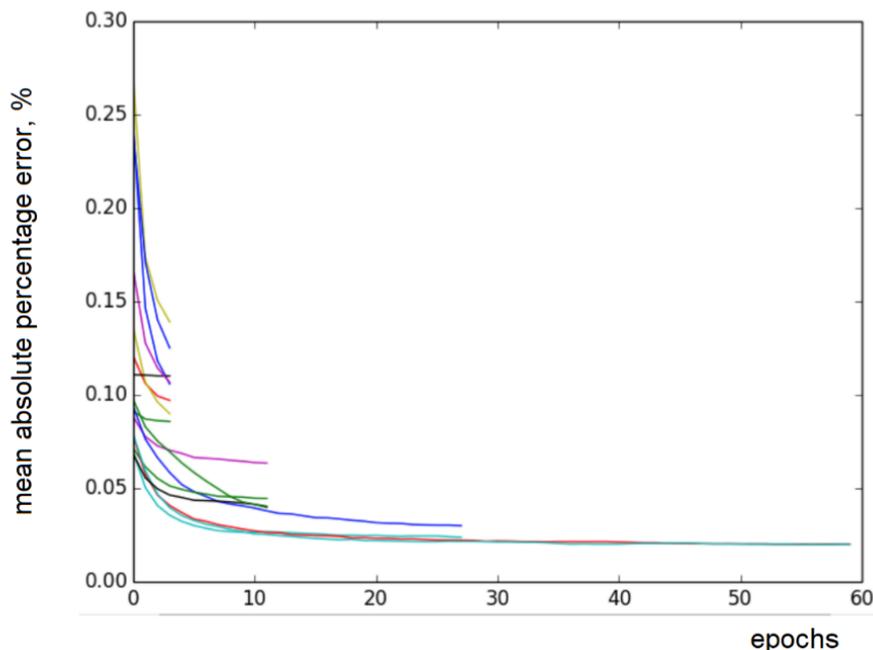


Рисунок 2 - Кривые обучения нейросетевых моделей при работе алгоритма Hyperband

Для увеличения ассоциативной способности, и, как следствие, увеличения точности STLF ГТП ГП, вышеперечисленные нейросети объединились в ансамбль. Структурная схема ансамбля нейросетей, предназначенных для краткосрочного прогнозирования почасового электропотребления ГТП ГП представлена на рисунке 3.

Данный ансамбль нейросетей имеет более 220 тыс. свободных параметров. Данная нейронная сеть, ввиду большого количества свободных параметров, будет склонна к переобучению. Стандартные методы борьбы с переобучением, такие как случайное прореживание связей между нейронами (Dropout) [9] и пакетная нормализация входных данных (Batchnormalization) [10], малоприменимы при прогнозировании временных рядов, так как нарушают внутреннюю структуру данных, обусловленную фактором времени. Поэтому, единственным способом борьбы с переобучением нейросети при STLF ГТП ГП является ранняя остановка алгоритма обучения. Данный метод борьбы с переобучением показывает высокую эффективность в следствии того, что затруднительно подобрать необходимое количество итераций обучающего цикла. В следствии того, что начальные значения весов синаптических связей нейросети имеет случайное распределение по нормальному закону, то практически невозможно определить «ландшафт» поверхности ошибки. Поэтому, неизвестно за какое количество эпох обучения рабочая точка алгоритма обучения достигнет до глобального минимума поверхности ошибки. Соответственно, возникает необходимость остановить алгоритм обучения после достижения окрестности глобального ми-

нимума. Показателем достижения рабочей точкой алгоритма обучения окрестности глобального минимума на поверхности ошибки является факт малого изменения абсолютной ошибки в течении эпохи обучения:

$$\frac{1}{C} \sum_{j \in C} \Delta e_j(n) \leq 0.005 - 0.01, \quad (11)$$

где $\Delta e_j(n)$ – изменение абсолютной ошибки нейрона j выходного слоя в течении эпохи обучения n .

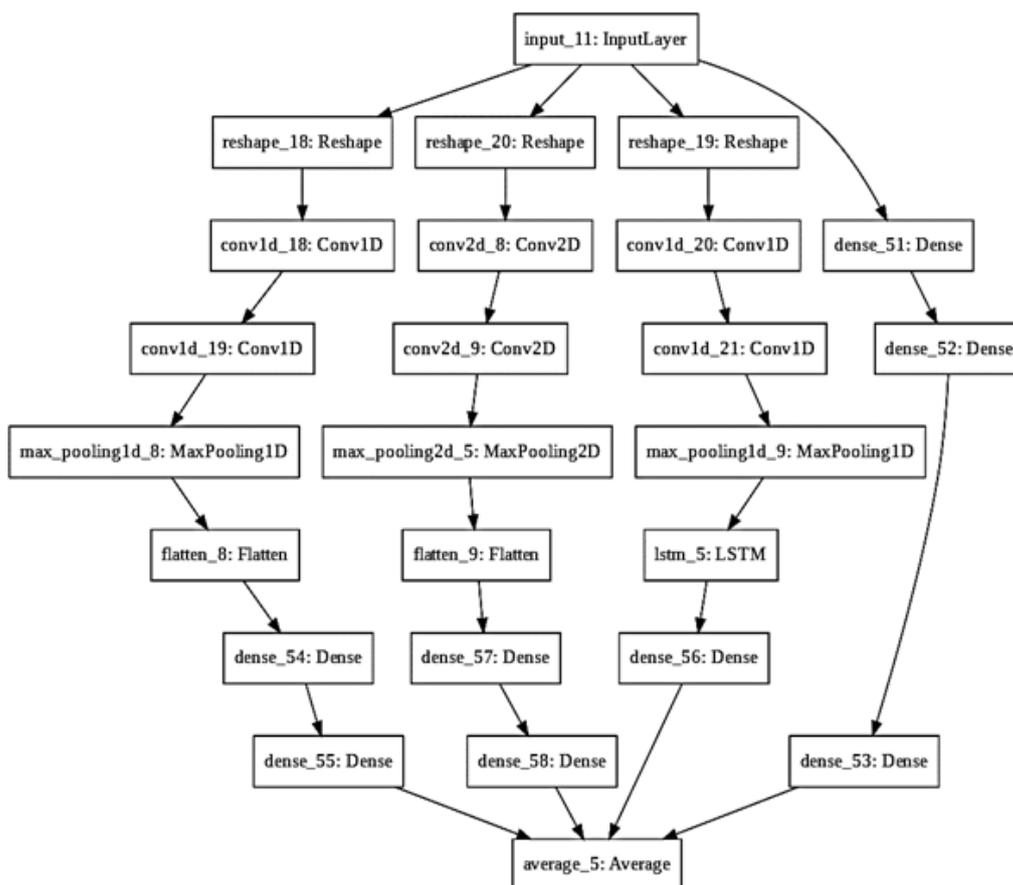


Рисунок 3 - Структурная схема ансамбля нейросетей, предназначенных для краткосрочного прогнозирования почасового электропотребления ГТП ГП

В ходе данного исследования произведен сравнительный анализ точности краткосрочного прогнозирования почасового электропотребления ГТП ГП, полученного с помощью метода экспертных оценок, многослойного персептрона (MLP), одномерной сверточной нейросети (1D_CNN), трехмерной сверточной нейросети (3D_CNN), рекуррентной нейросети типа долго-краткосрочной памяти (LSTM) и ансамблевого нейросетевого алгоритма (Ensemble). На рисунке 4 представлены кривые изменения средней абсолютной процентной ошибки прогнозирования нейросетевых алгоритмов.

Применение ансамблевого нейросетевого метода для краткосрочного прогнозирования почасового электропотребления ЭТК РЭС позволило снизить

относительную ошибку прогнозирования на тестовой выборке на 1,25 % по сравнению с прогнозом, полученным с помощью трехслойного персептрона.

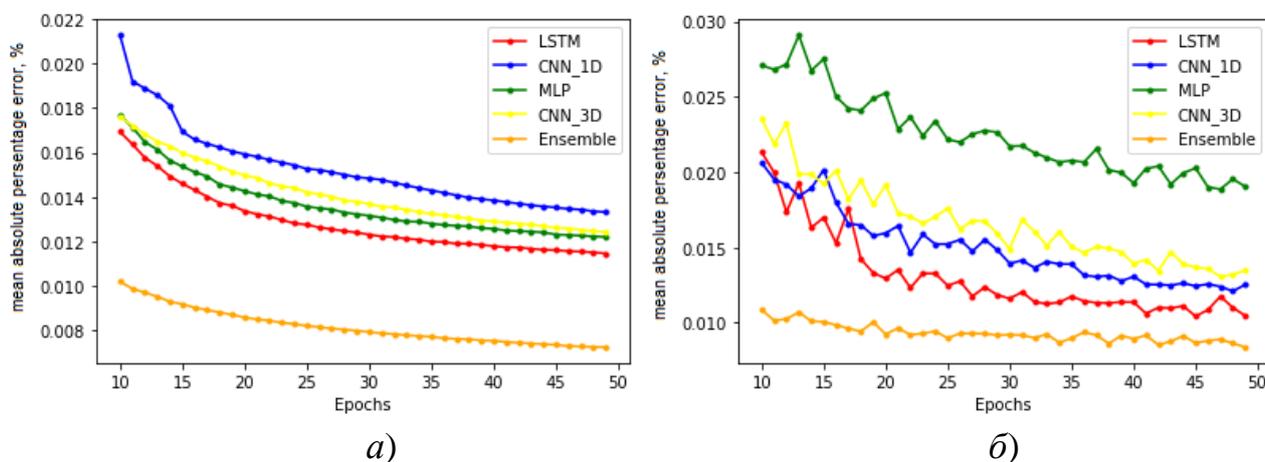


Рисунок 4 - Кривые изменения средней абсолютной процентной ошибки прогнозирования нейросетевых алгоритмов: а) тренировочной выборке данных; б) тестовой выборке данных.

Выбор оптимальной архитектуры, конфигурации, а также алгоритма обучения нейросетевой модели играет ключевую роль при краткосрочном прогнозировании потребления электроэнергии с помощью инструментов машинного обучения. Применение ансамбля ИНС, обучаемого с помощью алгоритма адаптивной инерции, для краткосрочного прогнозирования почасового электропотребления группы точек поставки гарантирующего поставщика электроэнергии позволило добиться снижения средней абсолютной процентной ошибки прогнозирования на ретроспективных 1,25 % по сравнению с прогнозом, полученным с помощью трехслойного персептрона.

Список используемой литературы

1. Доманов В. И., Билалова А. И. Прогнозирование объемов энергопотребления в зависимости от исходной информации // Вестник ЮУрГУ. Серия «Энергетика». 2016. Т. 16. № 2. С. 59–65. DOI: 10.14529/power160208
2. Katasev A. S. Neuro-fuzzy model of fuzzy rules formation for objects state evaluation in conditions of uncertainty // Computer Research and Modeling. 2019. Vol. 11 (3). P. 477–492. DOI: 10.20537/2076-7633-2019-11-3-477-492.
3. Станкевич Т. С. Разработка метода оперативного прогнозирования динамики развития лесного пожара посредством искусственного интеллекта и глубокого машинного обучения // Вестник Иркутского государственного технического университета. – 2018. – Т. 22. – № 9. – С. 111–120. DOI: 10.21285/1814-3520-2018-9-111-120.
4. Masood N. A., Ahsan Q. A Methodology for identification of weather sensitive component of electrical load using empirical mode decomposition technique // Energy and Power Engineering. 2013. Vol. 5. P. 293-300.
5. Хомутов С. О., Серебряков Н. А. Создание нейросетевой математической модели краткосрочного прогнозирования электропотребления электротехнического комплекса участка районных электрических сетей 6–35 кВ // Транспортные системы и технологии. 2020. Т. 6. № 1. С. 80–91. DOI: 10.17816/transsyst20206180-91.

6. Kingma D. P., Lei Ba J. ADAM: a method for stochastic optimization. URL: <https://arxiv.org/pdf/1412.6980.pdf> (дата обращения: 30.09.2023)
7. Воевода А. А., Романников Д. О. Синтез нейронной сети для реализации рекуррентного метода наименьших квадратов // Научный вестник НГТУ. 2018. № 3 (72). С. 33–42. doi: 10.17212/1814-1196-2018-3-33-42.
8. Li L., Jamieson K., Desalvo G., Rostamizadeh A., Talwalkar A. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. URL: <https://arxiv.org/pdf/1603.06560.pdf> (дата обращения: 18.10.2023).
9. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting // Journal of Machine Learning Research. 2014. Vol. 15, P. 1929-1958.
10. Ioffe S., Szegedy C. Batch Normalization: Accelerating deep network training by reducing internal covariate shift. URL: <https://arxiv.org/pdf/1502.03167.pdf> (дата обращения: 02.10.2023).

Информация об авторах

Серебряков Н. А. – к.т.н., доцент, ФГБОУ ВО «Алтайский государственный технический университет им. И. И. Ползунова», РФ, Алтайский край, г. Барнаул.

Ссылка для цитирования

Серебряков, Н. А. Выбор оптимальной архитектуры и конфигурации нейросетевой модели при краткосрочном прогнозировании электропотребления гарантирующих поставщиков электроэнергии / Н. А. Серебряков // Энерджинет. 2023. № 1. URL: <http://nopak.ru/231-705> (дата обращения: 25.10.2023).

