

Robust Short-Term Wind Speed Forecasting Using Adaptive Shallow Neural Networks

Matrenin P.V., Manusov V.Z., Igumnova E.A.

Novosibirsk State Technical University
Novosibirsk, Russian Federation

Abstract. Wind speed forecasting is necessary to integrate wind farms into power systems. In the past ten years, the forecasting models have become increasingly complex due to the development of artificial intelligence methods and computing power. Simultaneously, the robustness of models has decreased since complex models have a high risk of overfitting and decline in the accuracy if working conditions change significantly. This work aims to develop a machine learning model for short-term wind speed forecasting with acceptable accuracy but high robustness and the possibility of automatic online retraining. A shallow multilayer perceptron, trained only on retrospective data on wind speed, is proposed. The most significant results are combining simple neural network architecture with ReLU activation function, Adam training method developed for deep neural networks; and the automatic hyper-parameters selection using Grid search with open upper bounds. The model was trained on the data of the autumn period and tested on the winter data. A comparison was made with the simplest and most robust adaptive forecasting methods: Brown and Holt models. The significance of the obtained results is that shallow neural networks using ReLU, Adam, and Grid search are practically not inferior to adaptive models in terms of tuning speed and the risk of subsequent differences in accuracy between training data and data supplied during operation. At the same time, shallow neural networks make it possible to obtain more accurate forecasts, and due to their small size, they are trained quickly; and retraining can be performed automatically when new data arrives.

Keywords: short-term forecasting, wind energy, adaptive methods, shallow neural network.

DOI: 10.5281/zenodo.4018960

UDC: 620.91

Predicție durabilă a vitezei vântului pe termen scurt utilizând rețele neuronale compacte adaptive

Matrenin P.V., Manusov V.Z., Igumnova E.A.

Universitatea Tehnică de Stat din Novosibirsk
Novosibirsk, Federația Rusă

Rezumat. Pronoza vitezei vântului este esențială pentru integrarea parcurilor eoliene în sistemele electroenergetice. În ultimii 10 ani, odată cu dezvoltarea metodelor de inteligență artificială și a puterii de calcul, modelele de prognoză au devenit din ce în ce mai complexe. Scopul lucrării este de a dezvolta un model de învățare automată pentru prognoza pe termen scurt a vitezei vântului cu o precizie acceptabilă cu capacități ridicate de stabilitate în timpul funcționării ulterioare și cu posibilitatea de formare suplimentară automată în timp real. Pentru a atinge acest obiectiv, se propune un perceptron compact multistrat, care poate fi instruit numai cu privire la datele retrospective ale vitezei vântului. Cele mai semnificative rezultate costau în combinația unei arhitecturi simple a modelului rețelei neuronale cu funcția de activare ReLU, metoda de antrenament Adam dezvoltată pentru rețelele neuronale profunde, precum și în procedura de reglare automată a hiper-parametrilor rețelei utilizând metoda de căutare S-a realizat analiză comparativă cu cele mai simple și mai fiabile metode de prognoză adaptivă: modelele Brown și Holt. Semnificația studiului constă în confirmarea obținută că rețelele neuronale compacte atunci când se utilizează căutarea ReLU, Adam și Grid nu sunt practic inferioare modelelor adaptive în ceea ce privește viteza de reglare a sarcinii și riscul unei discrepante ulterioare în ceea ce privește precizia datelor de antrenament și a datelor care vor fi furnizate în timpul funcționării. În același timp, rețelele neuronale compacte permit obținerea de previziuni mai precise și, datorită dimensiunilor mici, sunt instruite rapid, iar antrenamentul poate fi efectuat automat la sosirea de noi date.

Cuvinte-cheie: prognoză pe termen scurt, energie eoliană, metode adaptative, rețea neuronală compactă.

Устойчивое краткосрочное прогнозирование скорости ветра с помощью адаптивных компактных нейронных сетей

П.В. Матренин, В.З. Манусов, Е.А. Игумнова

Новосибирский государственный технический университет
Новосибирск, Российская Федерация

Аннотация. Прогнозирование скорости ветрового потока необходимо для интеграции ветровых электростанций в электроэнергетические системы. В последние 10 лет с развитием методов искусственного интеллекта и вычислительных мощностей модели прогнозирования становятся все более

сложными. Но при этом теряется устойчивость таких моделей к изменениям условий работы, так как сложные модели имеют высокий риск переобучения. Целью работы является разработка модели машинного обучения для краткосрочного прогнозирования скорости ветра с приемлемой точностью, но высокой устойчивостью при последующей эксплуатации, и с возможностью автоматического дообучения в реальном времени. Для достижения поставленной цели предложен компактный многослойный перцептрон, обучаемый только на ретроспективных данных о скорости ветра. Наиболее существенными результатами являются соединение простой архитектуры нейросетевой модели с функцией активации ReLU, методом обучения Adam, разработанными для глубоких нейронных сетей, а также в процедуре автоматической настройки гипер-параметров сети с помощью метода Grid search с открытыми верхними границами. Модель обучалась на данных осеннего периода, а тестировались на данных зимнего периода, таким образом исследование было приближено к реальной ситуации, когда разработанная модель вводится в эксплуатацию и должна без внесения изменений начать работать в новых условиях. Было проведено сравнение с наиболее простыми и надежными адаптивными методами прогнозирования: моделями Брауна и Хольта. Значимость исследования заключается в полученном подтверждении того, что компактные нейронные сети при использовании ReLU, Adam и Grid search практически не уступают адаптивным моделям с точки зрения скорости настройки на задачу и риска последующего расхождения точности на обучающих данных и данных, которые будут подаваться во время эксплуатации. При этом компактные нейронные сети позволяют получать более точные прогнозы, а за счет малого размера они быстро обучаются и обучение может выполняться автоматически при поступлении новых данных.

Ключевые слова: краткосрочное прогнозирование, ветроэнергетика, адаптивные методы, компактная нейронная сеть.

ВВЕДЕНИЕ

Энергия ветра относится к наиболее важным источникам возобновляемой энергии, а ветроэнергетика последние годы набирает все большую популярность как дополнительный источник электроэнергии для систем электроснабжения стран и регионов, а также для отдельных домохозяйств [1, 2], в 2019 году суммарная мощность ветровых электростанций (ВЭС) оценивается в 650 ГВт [2]. Энергия, получаемая от ветра, является трудно предсказуемой, что осложняет интеграцию ВЭС в электроэнергетические системы [3, 4]. Даже краткосрочное прогнозирование скорости ветра представляет собой сложную задачу, поскольку изменение ветра даже в ближайший час зависит от многих быстро меняющихся факторов. Поэтому можно говорить о высокой актуальности исследования методов, позволяющих предсказывать скорость ветра хотя бы на короткий промежуток времени. В конечном счете прогнозировать важнее не скорость ветра, а вырабатываемую мощность, но в данной работе прогноз делается именно для скорости, так как вырабатываемая мощность является производной величиной и зависит от характеристик ВЭС.

Можно выделить несколько основных групп методов, которые используются для прогнозирования скорости ветра [5]: 1) построение физических метеорологических

моделей; 2) статистические методы; 3) методы искусственного интеллекта (машинного обучения).

Первая группа моделей использует большое число метеорологических данных и сложные модели движения атмосферы [6].

В качестве примеров можно выделить работу, использующую модель на основе фильтра Калмана [7], и исследование [8], в котором изменение метеорологических параметров описывается с помощью Гауссова процесса.

Такой подход может обеспечить необходимую точность прогноза [3, 6], но предъявляет высокие требования к точности входных метеорологических данных, нуждается в архиве наблюдений за длительный период времени и отличается высокой вычислительной сложностью.

Статистический подход, напротив, намного проще и в общем случае может быть применен при наличии ретроспективных данных только о скорости ветрах. Как и в других областях, где необходим прогноз временного ряда, широко используются различные авторегрессионные методы, чаще других это модификации ARIMA (autoregressive integrated moving average) [9, 10] и методы, использующие экспоненциальное сглаживание [11]. Одним из главных преимуществ таких методов является низкая вычислительная сложность (приводит к простоте настройки, низкому риску ошибки в применении, низким

требованиям к вычислительным ресурсам) относительно физических моделей и методов искусственного интеллекта. Но их точность не всегда удовлетворительная. Для повышения точности авторы часто используют гибридные модели, объединяющие принципы авторегрессии в сочетании с некоторым методом фильтрации. В работе [12] к методу ARMA (auto-regressive and moving average model) добавлен фильтр Калмана, а в исследовании [13] ARMA соединяется с вейвлет-преобразованием. Обзор гибридных методов прогнозирования скорости ветра [14] выделяет такие направления гибридизации как предобработка признаков (pre-processing), выбор признаков с помощью методов оптимизации, обработка и корректировка ошибок (post-processing).

Недостатком гибридных методов является сверхаддитивное повышение сложности, что приводит к высокой трудоемкости настройки метода для конкретного места и снижению универсальности построенной модели. В результате теряется и главное указанное преимущество – простота.

Последняя группа методов – интеллектуальные модели – отличается очень высоким разнообразием. Можно утверждать, что к задаче прогнозирования скорости ветра были применены все разделы искусственного интеллекта, чаще других используются искусственные нейронные сети (ИНС) [15-21]; нечеткая логика [17, 21, 22]; метод опорных векторов (SVM) [17, 23]; метаэвристические популяционные методы оптимизации (эволюционные и роевые) [15, 16, 21, 24]. В таких исследованиях тоже часто используются гибридные модели: ELM (extreme learning machine) и MOGWO (multi-objective grey wolf optimization) [15], LSTM (Long-Short Term Memory) нейронная сеть, ELM и эволюционная оптимизация [16], нечеткие нейронные сети [17], вейвлет-нейронные сети [19]. Наиболее сложная комбинация среди приведенных работ предложена в исследовании [21]: нечеткая нейронная сеть соединена с оптимизационным алгоритмом роя частиц и вейвлет-преобразованием.

Как показано в обзорах [3, 6], методы искусственного интеллекта позволяют достичь высокой точности в лабораторных условиях, то есть, когда коллектив авторов с использованием мощных вычислительных

ресурсов создает модель для определенной выборки данных. Но при этом возникает проблема переобучения (overfitting) и снижения надежности, когда несмотря на кросс-валидацию и разделение на обучающую, валидационную и тестовую выборки, построенная модель все-таки оказывается подогнана к рассматриваемой в конкретном исследовании задаче. Из-за этого возникают все более сложные гибриды, поскольку авторы начинают делать акцент не на практической ценности, а на научной; на достижении минимума ошибки прогноза, а не на создании простого и надежного метода с низким риском переобучения.

В данной работе, напротив, рассмотрена возможность применения очень простых моделей: адаптивных моделей Брауна и Хольа и простейший вариант ИНС – многослойный перцептрона с небольшим числом слоев и нейронов. Отличие подхода в том, что задача рассматривается с позиции оценки – можно ли при проектировании ВЭС, не проводя большой исследовательской работы, и с небольшой выборкой данных построить модель прогнозирования приемлемого качества, которая после ввода в эксплуатацию будет работать с точностью, близкой к той, что получена на этапе построения модели. Кроме того, важно отметить, что с развитием систем накопления электроэнергии требования к точности прогноза снижаются, так как ошибки прогноза могут корректироваться путем накопления или извлечения электроэнергии из накопителя. Поэтому представляется актуальной задача повышения не столько точности, сколько простоты, надежности и устойчивости моделей к изменению условий работы ВЭС. Целью работы является разработка модели машинного обучения для краткосрочного прогнозирования скорости ветра с приемлемой точностью, но высокой устойчивостью при последующей эксплуатации, и с возможностью автоматического дообучения в реальном времени.

I. МЕТОДЫ ИССЛЕДОВАНИЯ

A. Выборка данных и показатели точности прогноза

В работе использована выборка почасовых значений скорости ветра на острове Русский за 2017 год, а именно данные за 2925 часов

(сентябрь-декабрь) [25]. Усредненные значения представлены в таблице 1. Фрагмент данных показан на рис. 1. Остров Русский представляется очень перспективным для строительства ВЭС [25]. Чтобы избежать подгонки моделей под данные, выборка была разделена, 2/3 (1950 часов) для обучения и настройки моделей, 1/3 (975 значений) для тестирования. В отличие от многих работ, выборка перед разделением не была перемешана, то есть модель обучалась на данных с сентября по середину ноября, а проверялась затем на данных с середины ноября по декабрь. Это очень важно, так как при перемешивании и последующем равномерном разделении

выборки на обучающую и тестовую в обе выборки попадают близкие данные. В результате формально соблюдаются требования по борьбе с переобучением, но по сути – нет, модель и обучается, и проверяется на одних и тех же данных. В настоящей работе в обучающую выборку и тестовую попали данные, разделенные по календарному признаку. Как это могло бы произойти в жизни, когда после построения модели она вводится в эксплуатацию и получает на вход новые данные. Через месяц эксплуатации можно будет обновить модель, дообучив на новых данных, но хотя бы месяц она должна проработать стабильно.

Таблица 1¹

Среднемесячные скорости ветра о. Русский²

Время суток, ч., Time of the day, h.	Сентябрь, м/с September	Октябрь, м/с October	Ноябрь, м/с November	Декабрь, м/с December
0	9,5	11,3	11,9	10,8
3	10,2	11,6	11,2	10,7
6	9,3	11,6	11,4	11,2
9	9,9	11,3	10,6	10,8
12	10,2	10,7	10,4	11,5
15	10,9	10,5	10,5	11,4
18	10,2	10,8	10,9	10,6
21	9,7	11,0	11,1	10,5

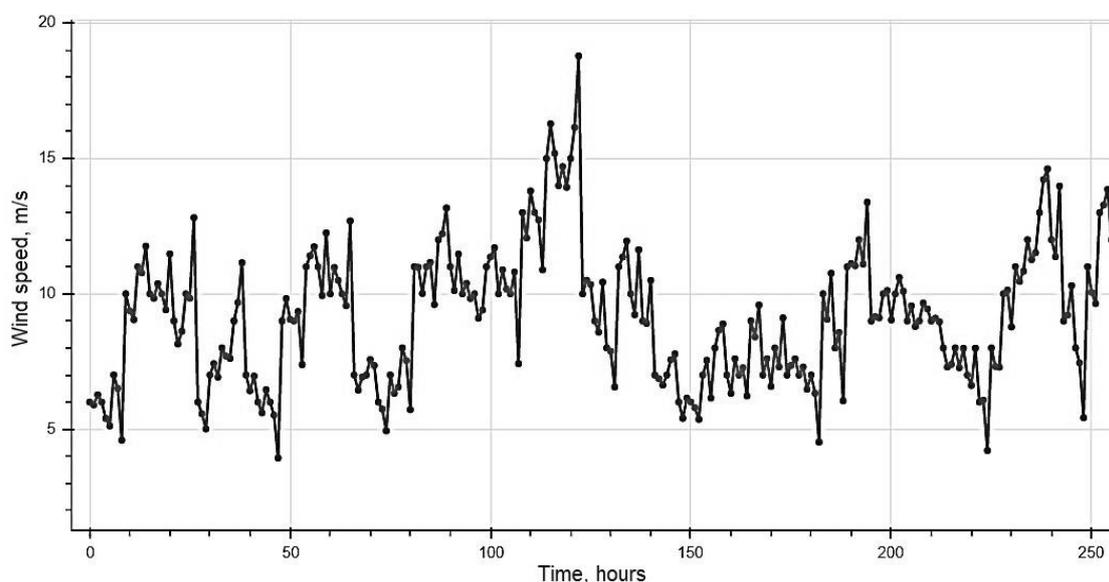


Рис.1. Фрагмент данных о скорости ветра.³

Математически задача сводится к снижению разницы между прогнозом и истинным значением скорости ветра. Чаще всего в исследованиях прогноза скорости ветра используют RMSE (Root Mean Squared Error) и MAPE (Mean Absolute Percentage

Error) [6].

В данной работе, так как используются только ретроспективные данные о скорости ветра, показатели можно записать следующим образом:

^{1,2,3} Appendix 1

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*(y_{[i-w:i-1]}))^2} \quad (1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y_i^*(y_{[i-w:i-1]})|}{y_i} \cdot 100\% \quad (2)$$

где y – истинное значение скорости ветра в i -й час; y^* – прогнозируемое значение на i -й час; $y_{[i-w:i-1]}$ – вектор истинных почасовых значений скорости за предыдущие w часов; n – размер выборки.

В качестве основного показателя в данной работе использован средний модуль относительной ошибки в процентах (MAPE), он менее восприимчив к большим ошибкам, которые неизбежно возникают при прогнозе ветра, так как скорость ветра в отдельные часы может меняться очень сильно, и эти изменения не будут укладываться в модель. Кроме того, для очень больших скоростей ветра ошибка оказывается не важна, поскольку ветроэнергетические установки всегда имеют ограничение сверху по вырабатываемой мощности. Проблемой использования MAPE является очень большие относительные ошибки при близких к нулю значениях истинной величины. Но климатические условия о. Русский таковы, что скорость ветра на часовом интервале никогда не приближается к нулю, поэтому MAPE подходит в качестве основной метрики точности. RMSE используется для наглядности, чтобы получать значение ошибки прогноза в физических единицах (м/с).

В. Адаптивные методы

Адаптивные методы основаны на моделях, которые приспособляются к изменениям, происходящим в рассматриваемом объекте или окружающей его среде [26]. Иными словами, адаптивная модель приспособляется к изменениям временного ряда, представляющего стохастический процесс.

Главное преимущество адаптивных моделей прогнозирования заключается в их способности быстро менять свое внутреннее состояние, реагируя на изменения в прогнозируемом объекте. В работе рассмотрены модели Брауна [27] и Хольта [28].

Модель Брауна может быть записана следующим образом:

$$\begin{aligned} S_0 &= y_0 \\ S_i &= \alpha y_{i-1} + (1-\alpha)S_{i-1}, i = 1, \dots, n \\ y_i^* &= S_i, i = 1, \dots, n \end{aligned} \quad (3)$$

α – параметр сглаживания, его значение задается от 0 до 1.

Модель Хольта немного сложнее:

$$\begin{aligned} S_1 &= y_1 \\ b_1 &= y_1 - y_0 \\ S_i &= \alpha y_{i-1} + (1-\alpha)(S_{i-1} + b_{i-1}), i = 2, \dots, n \\ b_i &= b_{i-1} + (S_i - S_{i-1}) + (1-b_{i-1})(S_{i-1} + b_{i-1}), i = 2, \dots, n \\ y_i^* &= S_i, i = 2, \dots, n \end{aligned} \quad (4)$$

α, β – параметры модели, их значения задаются от 0 до 1.

Выражения (3), (4) наглядно показывают простоту моделей Брауна и Хольта для реализации, отладки, применения; а также из них очевидна невозможность подгонки (переобучения), поскольку в них всего лишь 1-2 параметра, которые нужно подобрать. При этом для прогноза на 1 шаг вперед эти модели могут обеспечить приемлемую точность. Кроме того, такие модели не требуют большой выборки данных для настройки.

С. Искусственные нейронные сети

Искусственная нейронная сеть представляет собой математическую модель, основанную на очень упрощенных принципах работы биологических нейронных сетей. ИНС за счет нелинейного соединения результатов работы простых сумматоров (нейронов) позволяет аппроксимировать сложные математические зависимости. Нелинейность достигается за счет использования на выходе нейронов функций активации.

В данной работе для ИНС использован указанный выше подход применения простых моделей. Для ИНС простота может быть оценена общим числом обучаемых параметров, то есть весов и смещений нейронов.

В работе не применены сложные глубокие (deep), рекуррентные (recurrent) или сверточные (convolution) сети, напротив, использована простейшая архитектура многослойного перцептрона с очень небольшим числом нейронов в каждом слое.

Но при этом использованы современные функция активации ReLU и метод обучения Adam, которые были разработаны для глубоких и сверточных нейронных сетей.

II. РЕЗУЛЬТАТЫ ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ

Вычислительные эксперименты были проведены с использованием языка программирования Python 3 в среде разработки Google Colab. Для обучения нейронных сетей применялась библиотека машинного обучения TensorFlow.

А. Подбор параметров адаптивных методов

В методе экспоненциального сглаживания Брауна параметром выступает постоянная сглаживания α , которая задает скорость

реакции модели на изменения и в то же время влияет на способность модели сглаживать случайные отклонения. При построении модели Хольта нужно настроить параметры α и β , которые совместно определяют скорость реакции и чувствительность к отклонениям.

В вычислительных экспериментах использован метод Grid search, когда проверяются все варианты значений параметров из заранее выбранного списка. Если параметров несколько, проверяются все их комбинации. Результаты настройки адаптивных моделей представлены в таблицах 2-3. В таблице 3 для компактности приведены не все варианты, а только наилучшие. Полученные наилучшие варианты применены на тестовой выборке, результаты показаны в таблице 4.

Таблица 2⁴

Настройка модели Брауна на обучающей выборке⁵

α	RMSE, м/с	MAPE, %
0.1	2.77	22.55
0.2	2.52	20.29
0.3	2.39	18.90
0.4	2.32	17.99
0.5	2.29	17.34
0.6	2.29	16.80
0.7	2.30	16.45
0.8	2.34	16.19
0.9	2.39	16.11
0.95	2.35	16.16

Таблица 3⁶

Настройка модели Хольта на обучающей выборке⁷

α	β	RMSE, м/с	MAPE, %
0.1	0.1	3.10	25.38
0.2	0.1	2.69	21.57
0.3	0.1	2.49	19.50
0.4	0.1	2.39	18.36
0.5	0.1	2.30	17.59
0.6	0.1	2.32	17.00
0.6	0.2	2.33	17.12
0.6	0.3	2.35	17.23
0.6	0.4	2.36	17.34
0.7	0.1	2.32	16.55
0.8	0.1	3.12	23.76

Таблица 4⁸

Результаты моделей Хольта и Брауна на обучающей и тестовой выборках⁹

Модель	α	β	RMSE обучение, м/с	MAPE обучение, %	RMSE тест, м/с	MAPE тест, %
Брауна	0.7	-	2.30	16.45	2.21	16.18
Хольта	0.7	0.1	2.32	16.55	2.24	16.35

В. Выбор числа слоев и нейронов ИНС

В обучении ИНС в качестве метрики для сравнения эффективности различных архитектур и настроек использован показатель MAPE, поскольку он менее подвержен влиянию отдельных сильных отклонений. В то же время в процессе обучения использован MSE (RMSE²), поскольку квадратичная функция потерь предпочтительнее для обучения с помощью методов градиентного спуска.

Для ИНС также, как и для адаптивных моделей использован метод Grid search, но с открытой верхней границей.

Количество скрытых слоев в работе изменялось от двух и до тех пор, пока не возник эффект переобучения, который хорошо заметен по расхождению точности на обучающей и тестовой выборках. По тому же

принципу менялось количество нейронов. Для упрощения числа экспериментов было решено использовать в каждом внутреннем слое одинаковое число нейронов. Данные приведены в таблице 5.

Наилучшая конфигурация имеет три скрытых слоя по 16 нейронов, при этом RMSE на обучающей выборке 1.29 м/с, на тестовой – 1.37 м/с. Кроме того, изначально на вход подавались данные за 120 часов (5 суток), но в ходе экспериментов оказалось, что достаточно 12 часов.

С. Выбор функции активации и метода обучения

Обучение нейронной сети, а далее и решение задач, зависит не только от количества скрытых слоев и нейронов на каждом слое, а также от функции активации и метода обучения.

Таблица 5¹⁰

Настройка числа слоев и числа нейронов¹¹

Скрытых слоев	Нейронов в слое	Обучающая выборка MAPE, %	Тестовая выборка MAPE, %
2	12	12.86	13.58
2	14	11.37	11.92
2	16	11.53	12.29
2	18	10.19	11.87
2	20	10.64	12.13
3	12	10.51	11.61
3	14	9.76	11.03
3	16	9.69	10.28
3	18	9.94	11.49
3	20	8.81	10.57
4	12	10.82	12.27
4	14	9.96	11.69
4	16	9.50	11.62
4	18	8.50	10.38
4	20	8.59	11.59
5	12	10.50	11.50
5	14	10.53	11.74
5	16	9.24	10.97
5	18	8.45	11.03
5	20	8.85	10.91

В работе проведено сравнение функций активации – сигмоидальной и ReLU, которая показывает высокую эффективность для глубоких сетей [29, 30], но может быть использована и для неглубоких сетей хотя бы

с одним скрытым слоем. Графики этих функций приведены на рис. 2, 3.

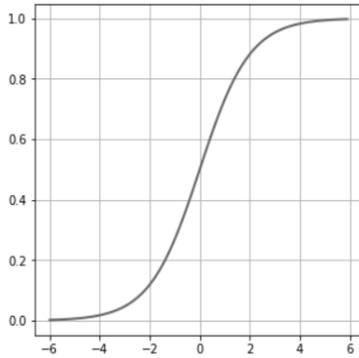


Рис.2. Функция активации Sigmoid.¹²

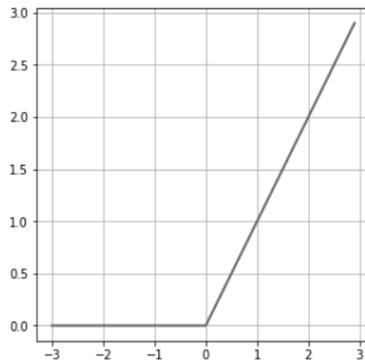


Рис.3. Функция активации ReLU.¹³

Из множества методов обучения были использованы классический стохастический градиентный спуск (Stochastic Gradient Descent, SGD) и модификация градиентного спуска Adam, который является комбинацией двух других методов: Momentum и RMSProp [31]. Его принципиальная схема может быть записана так:

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

где W – матрица весов;

V_{dw} – матрица, характеризующая инерционные свойства параметров ИНС, 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 (использовано значение 0.9 [31]);

S_{dw} – матрица, характеризующая степень (“энергию”, так как градиент возведен в квадрат) изменения параметров ИНС, без учета направления изменения;

β_2 – параметр, задающий баланс между учетом предыдущей энергии изменения направления градиента и направления градиента, полученного на очередной эпохе обучения и на очередном пакете (использовано значение 0.999 [31]);

ε – близкое к нулю положительное число для предотвращения деления на ноль (использовано значение $1E-6$ [31]);

α – величина шага обучения (использовано значение 0.01);

t – номер пакета.

Нужно отметить, что несмотря на популярность метода Adam для обучения глубоких сетей, ряд исследований показывает, что в общем случае этот метод хуже, чем SGD [32, 33]. Поэтому следует делать выбор метода обучения для каждой решаемой задачи, исходя из результатов вычислительных экспериментов.

В данной работе для разработанной архитектуры ИНС быстро было выявлено преимущество ReLU и Adam, поэтому в приведенной выше таблице 5 показаны результаты именно с их использованием. А в таблице 6 приведены результаты с использованием трех скрытых слоев по 16 нейронов. Рисунки 4 и 5 показывают фрагмент тестовой выборки с результатами прогноза с помощью модели Хольта и ИНС.

III. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Для достижения цели работы была предложена и исследована компактная ИНС, обладающая высокой устойчивостью к переобучению и высокой робастностью за счет малого числа слоев и нейронов. Чтобы подтвердить свойства разработанной модели, было выполнено сравнение с моделями Брауна и Хольта. Выборка данных была разделена на обучающую и тестовую, в тестовую попали данные за другой календарный период. На обучающей выборке были выполнены обучение ИНС и настройка параметров адаптивных моделей.

^{11,12} Appendix 1

Сравнение функций активации¹⁵

Метод обучения ИНС	Функция активации в нейронах скрытого слоя	Тестовая выборка MAPE, %	Обучающая выборка MAPE, %
SGD	Sigmoid	17.5	18.8
Adam	Sigmoid	15.5	16.2
Adam	ReLU	9.7	10.3
SGD	ReLU	15.4	15.6

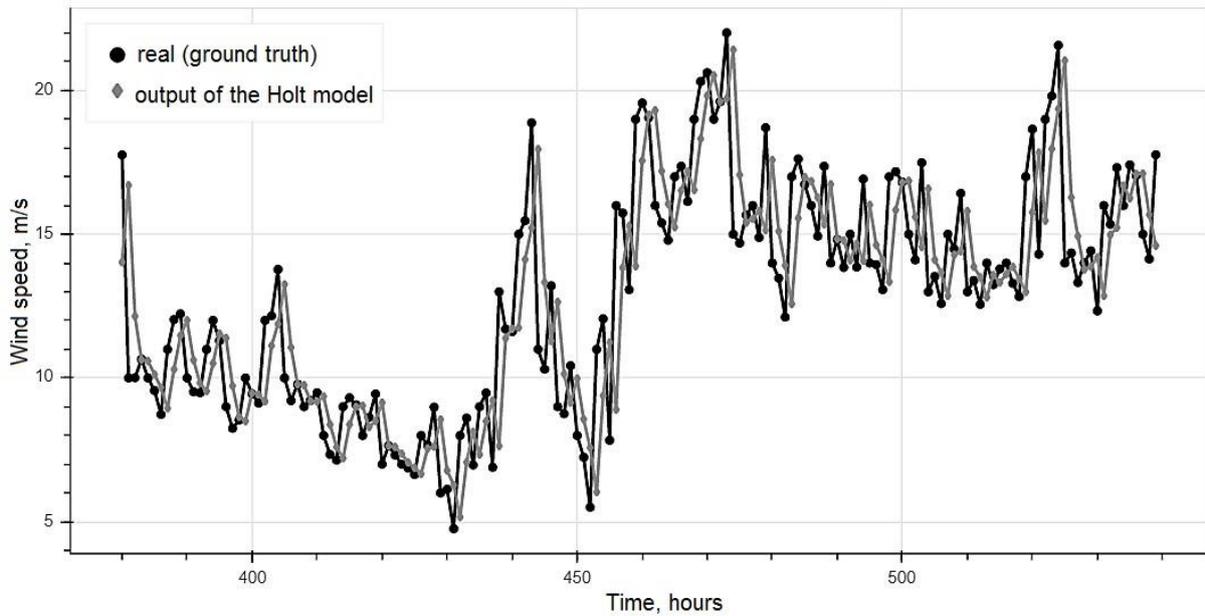


Рис. 4. Фрагмент данных с сопоставлением прогнозов модели Хольта и истинного результата¹⁶

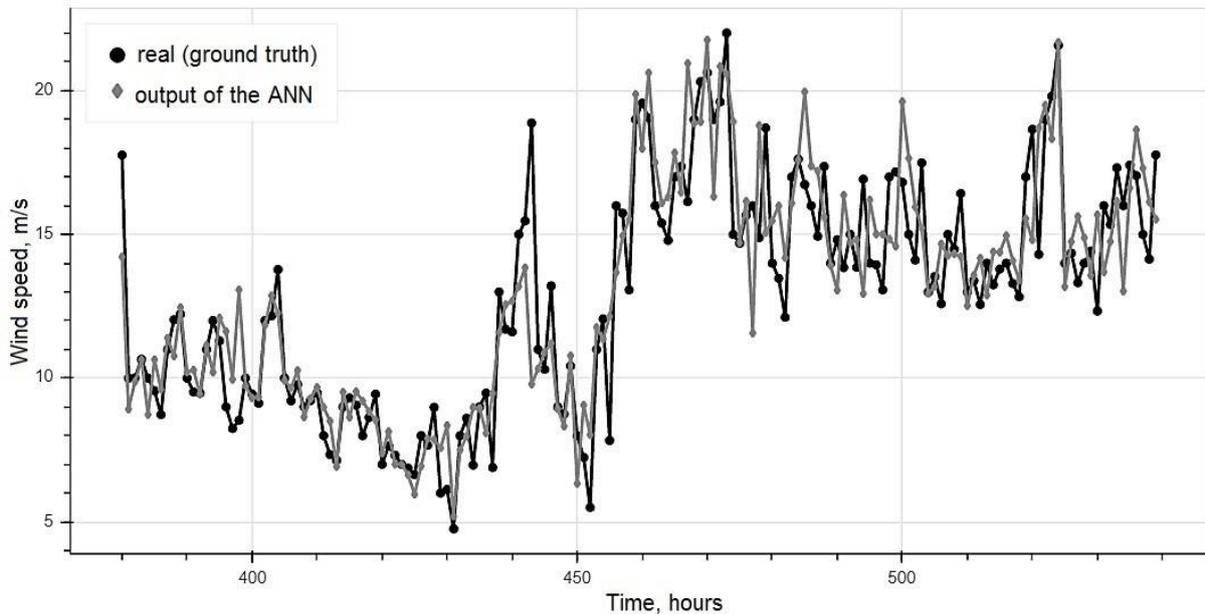


Рис. 5. Фрагмент данных с сопоставлением прогнозов ИНС и истинного результата¹⁷

Адаптивные модели Брауна и Хольта показали очень близкие друг к другу результаты (MAPE 16 %, RMSE 2.2 м/с) и отсутствие переобучения. Данные в таблице 3

показывают, что без подбора параметров точность адаптивных моделей может быть значительно ниже (на 5–10 пп.). Но благодаря простоте и отсутствию риска переобучения

они могут быть настроены на очень небольшой выборке и сразу быть введены в эксплуатацию.

Выбранная архитектура ИНС (многослойный перцептрон с малым числом нейронов в слоях) показала следующие особенности.

1. Не высокая, но приемлемая для задачи точность (MAPE 10 %, RMSE 1.4 м/с) на тестовой выборке, при этом намного выше, чем у адаптивных моделей (таблица 6).

2. Низкое переобучение: разница MAPE на обучающей и тестовой выборке составила 0.6 пп. Но при этом потребовалось применить подбор параметров с помощью Grid search, при слишком малом количестве слоев или нейронов точность существенно падала даже на обучающей выборке, при слишком большой уже начиналось переобучение (таблица 5).

3. Методы повышения точности и устойчивости к переобучению глубоких и сверточных ИНС, такие как Adam и ReLU, оказались полезными и для компактной ИНС, их применение существенно повысило точность на тестовой выборке (таблица 6).

4. Быстрое обучение: ИНС сходилась к решению за 1500-2500 эпох (4-6 минут на двухъядерном процессоре с частотой 2.2 ГГц), что позволяет сделать обучение автоматическим и перевести его в режим реального времени, то есть сделать ИНС адаптивной.

5. Модель имеет лишь два гипер-параметра, поэтому не требуется использовать недетерминированные методы настройки гипер-параметров, такие как Random search [34], которые по своей природе дают менее устойчивые результаты и также могут быть причиной снижения надежности модели.

Таким образом, архитектура компактного многослойного перцептрона позволила в рассматриваемой задаче краткосрочного прогнозирования скорости ветра использовать преимущества методов машинного обучения [16, 17], но без существенной потери качества с точки зрения скорости настройки и риска последующего расхождения точности на обучающих и эксплуатационных, то есть минимизировать риски применения машинного обучения для прогноза скорости ветра, приведенные в работах [3, 6].

Предложенная архитектура является многослойной, поэтому для нее эффективнее использовать ReLU и Adam, но благодаря малому числу нейронов в слоях не возникает необходимости применения дополнительных приемов борьбы с переобучением, указанных в [32]. Это позволяет оставить всего лишь 2 гипер-параметра: число скрытых слоев и число нейронов с скрытым слое, а для двух параметров можно применить детерминированный метод Grid search с высокой эффективностью [34]. В результате очень сильно упрощается подбор гипер-параметров, что также отличает предложенный подход от рассмотренных во введении работ, использующих методы машинного обучения.

ЗАКЛЮЧЕНИЕ

Проведено исследование компактных ИНС для краткосрочного прогнозирования скорости ветра о. Русский.

Рассматривалась не только точность прогноза, но и надежность (робастность) моделей и устойчивость к переобучению, то есть риск того, что после ввода в эксплуатацию точность прогноза окажется существенно ниже, чем полученная на этапе разработки. Было проведено сравнение предложенной модели компактного многослойного перцептрона с моделями Брауна и Хольа.

Предложенный подход позволяет достичь более высокой точности, чем адаптивные методы, и при этом практически без снижения простоты и робастности. Однако при этом необходимо использовать методы повышения эффективности обучения ИНС, такие как ReLU и Adam.

Предложенная модель благодаря низкому времени обучения и настройке гипер-параметров может быть быстро разработана для конкретных климатических условий с автоматическим подбором гипер-параметров и внедрена с очень малым риском существенного увеличения ошибки при реальной эксплуатации по сравнению с ошибкой, определенной на этапе разработки.

Кроме того, такая ИНС может в процессе эксплуатации периодически дообучаться на новых данных в автоматическом режиме и таким образом адаптироваться к изменениям метеорологических условий.

APPENDIX 1 (ПРИЛОЖЕНИЕ 1)

^{1,2}**Table. 1.** The average monthly wind speed of Russky Island.

³**Fig. 1.** Wind speed data snippet.

^{4,5}**Table.2.** Tuning the Brown model using the training set.

^{6,7}**Table.3.** Tuning the Holt model using the training set.

^{8,9}**Table.4.** Holt and Brown models' results.

^{10,11}**Table.5.** Selection the number of hidden layers and the number of neurons.

¹²**Fig. 2.** Sigmoid activation functions.

¹³**Fig. 3.** ReLU activation functions.

^{14,15}**Table.6.** Comparison of activation functions and training methods.

¹⁶**Fig. 4.** Outputs of Holt model and real wind speed.

¹⁷**Fig. 5.** Outputs of ANN model and real wind speed.

ACKNOWLEDGEMENTS

Исследование выполнено при финансовой поддержке в рамках реализации программы развития НГТУ, научный проект №С20-20.

Литература (References)

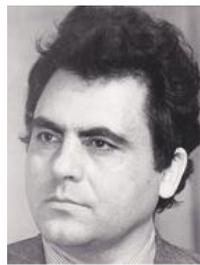
- [1] Niu T, Wang J., Lu H., and Du P. Uncertainty modeling for chaotic time series based on optimal multi-input multi-output architecture: Application to offshore wind speed. *Energy Conversion and Management*, 2018, vol. 156, pp. 597-617.
- [2] T. Wang. Installed wind power capacity - worldwide 2001-2019. 2019. Available at: <https://www.statista.com/statistics/268363/installed-wind-power-capacity-worldwide> (accessed 05.06.2020).
- [3] Zhang L., Dong Y., Wang J. Wind Speed Forecasting Using a Two-Stage Forecasting System with an Error Correcting and Nonlinear Ensemble Strategy. *IEEE Access*, 2019, vol. 7, pp. 176000-176023.
- [4] Wang Z., Zhang J., Zhang Y., Huang C., Wang L. Short-Term Wind Speed Forecasting Based on Information of Neighboring Wind Farms. *IEEE Access*, 2020, vol. 8, pp. 16760-16770.
- [5] Zhou Q., Wang C., Zhang G. Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems. *Applied Energy*, 2019, vol. 250, pp. 1559-1580.
- [6] Foley A.M., Leahy P.G., Marvuglia A., McKeogh E.J. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 2012, vol. 37, no. 1, pp. 1-8.
- [7] Cassola F., Burlando M. Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output. *Applied Energy*, 2012, vol. 99, pp. 154-166.
- [8] Fang S., Chiang H. A High-Accuracy Wind Power Forecasting Model. *IEEE Trans. on Power Systems*, 2017, vol. 32, no. 2, pp. 1589-1590.
- [9] Erdem E., Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy*, 2011, vol. 88, no. 4, pp. 1405-1414.
- [10] Wang J., Zhou Q., Zhang X. Wind power forecasting based on time series ARMA model. *IOP Conference Series, Earth Environment Science*, 2018, vol. 199, no. 2, Art. no. 022015.
- [11] Jonsson T., Pinson P., Nielsen H.A., Madsen H. Exponential smoothing approaches for prediction in real-time electricity markets. *Energies*, 2014, vol. 7, no. 6, pp. 3710-3732.
- [12] Cadenas E., Rivera W., Campos-Amezcuca R., Heard C. Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model. *Energies*, 2016, vol. 9, no. 109, 15 p.
- [13] Lei C., Ran L. Short-term wind speed forecasting model for wind farm based on wavelet decomposition. *Proc. 3rd Int. Conf. Electronic Utility Deregulation Restructuring Power Technology*, Nanjing, 2008, pp. 2525-2529.
- [14] Tascikaraoglu A., Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. *Renewable and Sustainable Energy Reviews.*, 2014, vol. 34, pp. 243-254.
- [15] Wu C., Wang J., Chen X., Du P., Yang W. A novel hybrid system based on multi-objective optimization for wind speed forecasting. *Renewable Energy*, 2020, vol. 146, pp. 149-165.
- [16] Hu Y.-L., Chen L. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm. *Energy Conversion and Management*, 2018, vol. 173, pp. 123-142.
- [17] Khosravi A., Koury R., Machado L., Pabon J., Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. *Sustainable Energy Technologies and Assessments*, 2018, vol. 25, pp. 146-160.
- [18] Yadav A.K., Malik H. Short-term wind speed forecasting for power generation in Hamirpur, Himachal Pradesh, India, using artificial neural networks. *Applications of Artificial Intelligence Techniques in Engineering*, Singapore, 2019, pp. 263-271.
- [19] Chitsaz H., Amjady N, Zareipour H. Wind power forecast using wavelet neural network trained by improved clonal selection algorithm. *Energy Conversion and Management*, 2015, vol. 89, pp. 588-598.
- [20] Quan H., Srinivasan D., Khosravi A. Short-term load and wind power forecasting using neural

- network-based prediction intervals. *IEEE Trans. of Neural Networks Learning Systems*, 2014, vol. 25, no. 2, pp. 303-315.
- [21] Osorio G., Matias J., Catalao J. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. *Renewable Energy*, 2015, vol. 75, pp. 301-307.
- [22] Jiang P., Yang H., and Heng J. A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting. *Applied Energy*, 2019, vol. 235, pp. 786-801.
- [23] Liu T., Jin Y., Gao Y. A new hybrid approach for short-term electric load forecasting applying support vector machine with ensemble empirical mode decomposition and whale optimization. *Energies*, 2019, vol. 12, no. 1520, 20 p.
- [24] Mirjalili S., Gandomi A.H., Mirjalili S.Z., Saremi S., Faris H., Mirjalili S. Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software*, 2017, vol. 114, pp. 163-191.
- [25] Khasanzoda N. *Optimizaciya rezhimov elektropotrebleniya v intellektual'nyh setyakh s dvustoronnim potokom energii metodami iskusstvennogo intellekta*. Diss. cand. tec. nauk [Optimization of power consumption modes in Swarm grid with two-way energy flow using artificial intelligence methods. Cand. tech. sci. diss.] Novosibirsk, 2019. 187 p.
- [26] Lukashin Y.P. *Adaptivnye metody kratkosrochnogo prognozirovaniya vremennykh ryadov* [Adaptive methods for short-term time series forecasting]. Moscow, 2003. 416 p.
- [27] Brown R.G. Smoothing Forecasting and Prediction of Discrete Time Series. New Jersey, 1963. 468 p.
- [28] C.C. Holt. Forecasting Trends and Seasonal by Exponentially Weighted Averages. *Int. Journal of Forecasting*. 2004, vol 20, no. 1 pp. 5-10.
- [29] Dahl G.E., Sainath T.N., Hinton G.E. Improving deep neural networks for LVCSR using rectified linear units and dropout. *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Vancouver, 2013, pp. 8609-8613.
- [30] LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, pp. 436-444.
- [31] Kingma D.P, Ba J. Adam: A method for stochastic optimization. 2014. Available at: <https://arxiv.org/abs/1412.6980> (accessed 15.05.2020).
- [32] Keskar N.S., Socher R. Improving Generalization Performance by Switching from Adam to SGD. 2017. Available at <https://arxiv.org/abs/07628v1> (accessed 18.05.2020).
- [33] Wilson A.C., Roelofs R., Stern M., Srebro N., Recht B. The Marginal Value of Adaptive Gradient Methods in Machine Learning. 2017. Available at: <https://arxiv.org/abs/1705.08292v2> (accessed 18.05.2020).
- [34] Bergstra J., Bengio, Y. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 2012, vol. 13, pp. 281-305)

Сведения об авторах.



Павел Викторович Матренин, кандидат технических наук. Закончил Новосибирский государственный технический университет (НГТУ) в 2014 г. Старший преподаватель кафедры систем электроснабжения предприятий НГТУ. Область научных интересов: системный анализ, оптимизация и управление в электроэнергетике.
E-mail: matrenin.2012@corp.nstu.ru



Вадим Зиновьевич Манусов, доктор технических наук, профессор. Закончил Новосибирский электротехнический институт в 1963 г. Профессор кафедры систем электроснабжения предприятий НГТУ. Область научных интересов: применение методов искусственного интеллекта в электроэнергетике.
E-mail: manusov@mail.ru



Евгения Александровна Игумнова, магистрант кафедры систем электроснабжения предприятий НГТУ. Область научных интересов: прогнозирование в электроэнергетике, возобновляемые источники энергии.
E-mail: e.igumnova97@gmail.com