

DOI: <https://doi.org/10.37162/2618-9631-2023-2-138-155>

УДК 556.536+519.688+004.8+004.438

Глубокие нейронные сети архитектуры трансформер в задачах гидрологических прогнозов

А.В. Романов¹, Э.Р. Акмаев², М.А. Червоненкис²

¹*Гидрометеорологический научно-исследовательский центр
Российской Федерации, г. Москва, Россия;*

²*ООО «ПАВЛИН Технологии» г. Дубна Московской обл., Россия
alexey.romanov@mecom.ru, akmaew1@gmail.com, chervonenkis@yandex.ru*

Выполнен теоретический анализ современных нейросетевых моделей, используемых для обработки временных рядов. При этом особое внимание уделено архитектуре построения алгоритмов глубокого машинного обучения. Показаны преимущества нейросетевой модели Temporal Fusion Transformer (TFT), которая выбрана в качестве базовой для моделирования процесса формирования весеннего половодья. Детально численно проанализированы возможности использования модели TFT для долгосрочного прогноза максимальных уровней воды с заблаговременностью 60 и 90 суток для нескольких водпостов р. Исеть (бассейн речной системы Тобола). В качестве исходной информации для обучения модели (зависимая/обучаемая выборка) использованы суточные временные ряды за 27 лет (1991–2017 гг.) по восьми гидрометеорологическим характеристикам. Приводятся результаты прогнозов для независимой выборки (2018–2022 гг.), а также данные оперативного прогноза за 2023 г. Выделены несколько направлений развития нейросетевого моделирования для долгосрочных и краткосрочных прогнозов речного стока.

Ключевые слова: долгосрочные гидрологические прогнозы, половодье, нейронные сети, расход воды, уровень воды, искусственный интеллект, глубокое машинное обучение, Временной Объединяющий Трансформер

Deep neural networks of transformer architecture in problems of hydrological forecasts

A.V. Romanov¹, E.R. Akmaev², M.A. Chervonenkis²

¹*Hydrometeorological Research Center of Russian Federation, Moscow, Russia*

²*PAWLIN Technologies, Dubna, Moscow region, Russia
alexey.romanov@mecom.ru, akmaew1@gmail.com, chervonenkis@yandex.ru*

The theoretical analysis of modern neural network models used for processing time series is carried out. At the same time, special attention is paid to the architecture of building deep machine learning algorithms. The advantages of the neural network model Temporal Fusion Transformer (TFT), which is selected as the base for modeling the process of spring flood formation, are shown. The possibilities of using the TFT model for a long-term forecast of maximum water levels with a lead time of 60 and 90 days for several points of the Iset river (the basin of the Tobol river system) are numerically analyzed in detail. The daily time series for 27 years (1991-2017) for eight hydrometeorological characteristics were used as the initial information for training the model (dependent (trainable) sample). The results of forecasts for an independent sample (2018-2022), as well as operational forecast data for 2023 are presented. Several directions of development of neural network modeling for long-term and short-term forecasts of streamflow are identified.

Keywords: hydrological long-term forecasts, flooding, neural networks, discharge, water level, water regime, artificial intelligence, deep machine learning, Temporal Fusion Transformer

Введение

Анализ мирового опыта развития методов математического моделирования речного стока показывает, что в основном все достигнутые в этом направлении результаты относятся к их использованию в области краткосрочных прогнозов водного режима с заблаговременностью от трех до десяти суток. При этом методы долгосрочных прогнозов с заблаговременностью месяц и более практически остались на уровне, достигнутом на рубеже XX и XXI веков. Физической основой последних прежде всего является уравнение водного баланса, которое может быть записано с той или точностью в зависимости от объема исходной гидрометеорологической информации. При этом подавляющее большинство методов долгосрочных прогнозов различных элементов речного стока, основанных на использовании уравнения водного баланса, базируются на физико-статистическом подходе, широко использующим корреляционные статистические методы анализа.

В то же время стремительный рост объема цифровой гидрометеорологической информации, развитие нейросетей, а также расширение парка и мощности вычислительных ресурсов предъявляют новые требования к их обработке, которые распространяются и на долгосрочные гидрологические прогнозы, представляющие собой практически чрезвычайно важную область гидрологии суши. На сегодняшний день в отечественной литературе практически нет публикаций, в которых рассматривается использование искусственных нейронных сетей (ANN) для долгосрочных прогнозов уровней воды с заблаговременностью месяц и более. В то же время практическая потребность в разработке таких методов постоянно нарастает, так как позволяет существенно повысить эффективность управления водными ресурсами, а также избежать серьезных экономических потерь при образовании катастрофических наводнений.

В последние годы алгоритмы глубокого обучения ANN показали большую эффективность в самых разных областях: от синтеза речи до компьютерного зрения и прогнозирования временных рядов [1]. Важной особенностью таких алгоритмов является их способность улавливать нелинейные связи между данными разного типа, а также работать с их большим объемом. В качестве объекта исследования возможностей использования ANN были выбраны три водпоста р. Исеть (бассейн речной системы р. Тобол), для которых классические методы долгосрочного прогноза максимальных уровней воды с заблаговременностью не менее 30 дней, используемые в ФГБУ «Уральское УГМС», не дали положительных результатов.

1. Архитектура базовой нейросетевой модели

Рассмотрим особенности архитектуры модели Temporal Fusion Transformer / Трансформатор Временного Синтеза (TFT) [7], которая нашла широкое применение в различных областях науки для предсказания временных рядов. Эта модель обладает множеством приложений и успешно используется в задачах прогнозирования временных данных, где эффективное предсказание и моделирование временных зависимостей играют важную роль [3]. Текущие часто используемые подходы включают в себя такие модели глубокого обучения, как MLP, RNN, LSTM [5], TCNN (Temporal Convolutional Neural Network / Временная Свёрточная Нейронная Сеть), а также комбинированные подходы, как, например, LSTM в сочетании со SWAT (Soil & Water Assessment Tool / Инструмент Оценки Почвы и Воды). Однако они все страдают одним и тем же недостатком: во многих случаях возможно использование только непрерывных временных наблюдаемых признаков, в то время как использование категориальных и статических признаков позволяет улучшить качество прогнозов. Подключение категориальных и статических признаков к указанным моделям возможно путем различных модификаций (например, используются one-hot – кодировщики, необучаемые слои или признаки подаются на вход верхних уровней модели). Однако их использование недостаточно эффективно (существенное увеличение размерности, неоптимальная архитектура).

В базовой модели TFT могут быть реализованы следующие признаки:

- Статические категориальные
- Статические непрерывные
- Временные категориальные, известные заранее
- Временные непрерывные, известные заранее
- Временные категориальные, наблюдаемые
- Временные непрерывные, наблюдаемые
- Целевые значения временных рядов, наблюдаемые до настоящего момента времени

Ещё одной проблемой, с которой сталкиваются модели, является не только необходимость обработки значений в различных диапазонах, но и способность предсказывать значения в достаточно большом интервале. Решением данной задачи может быть преобразование входного временного ряда целевой переменной к нормализованному распределению через среднее и стандартное отклонение. Эти значения могут использоваться для преобразования выхода модели в требуемый диапазон. Такой подход позволяет улучшить точность прогнозов и снизить ошибки модели [2], поскольку теперь её непосредственные выходы будут лежать в одном распределении.

Общая архитектура используемой модели представлена на рис. 1.

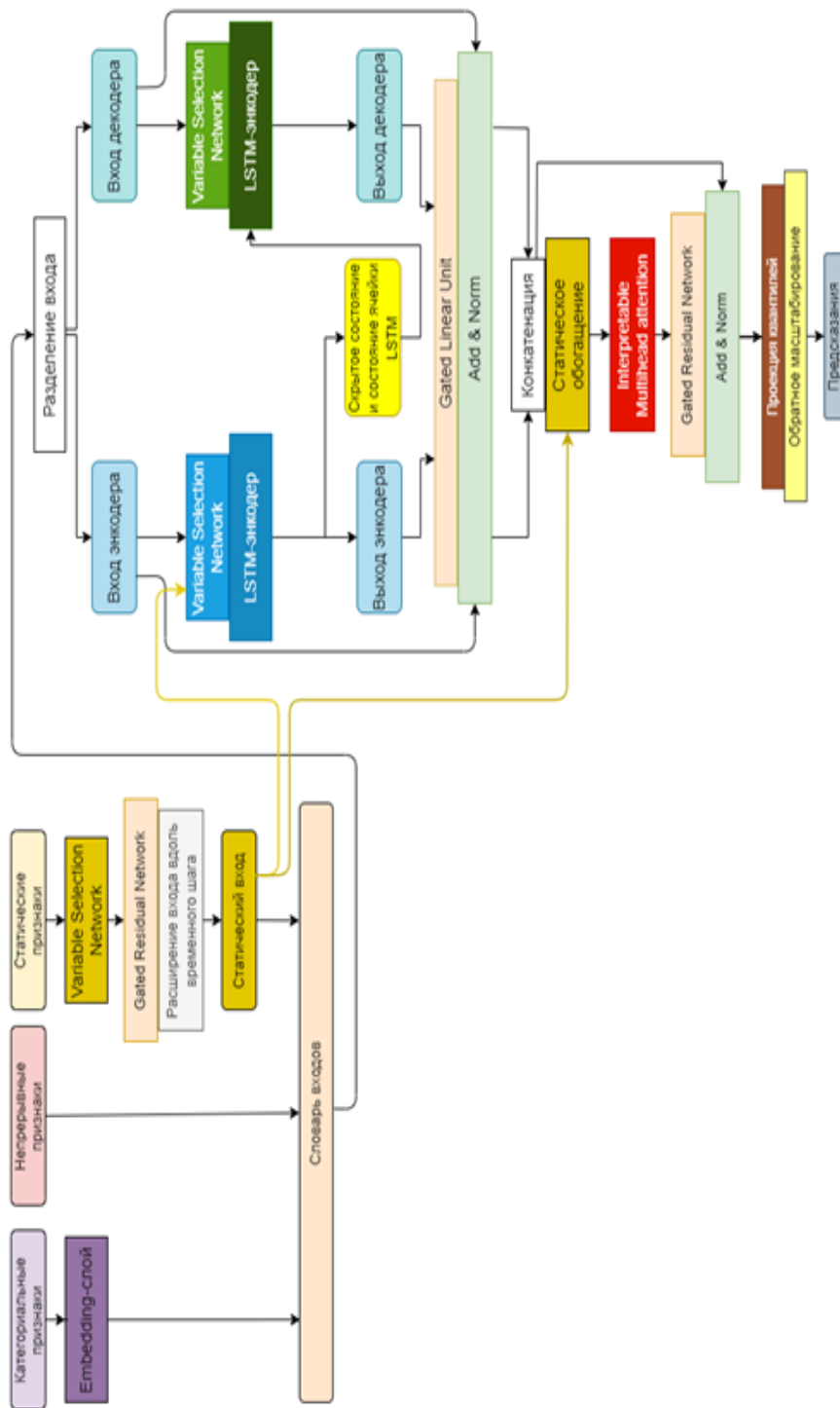


Рис. 1. Архитектура модифицированной модели Temporal Fusion Transformer.
 Fig. 1. Modified Temporal Fusion Transformer model architecture.

Выделим основные концептуальные идеи, положенные в основу построения данной модели:

- в качестве энкодера и декодера используется LSTM-слой;
- GLU-блоки или их включающие используются для контроля потока информации и предотвращения насыщения или затирания градиентов линейных слоев;
- сети выбора переменных (Variable Selection Networks) используются для отбора наиболее важных признаков, возвращая дополнительно сами веса важности признаков. Также этот блок ответственен за приведение размерности от разных типов данных к одному виду – например, категориальные признаки будут иметь отличную от непрерывных размерность;
- в качестве внимания используется скалярное внимание (dot-product attention).

В данной работе все численные эксперименты выполнены на базе модификации модели, предложенной в [6]. Для упрощения описания блоков используемой модели выделим только те из них, которые отличаются от оригинальной имплементации.

Регуляционные (Шлюзовые) линейные ячейки (Gated Linear Unit, GLU) – блоки регуляции потока информации в нейронных сетях [4]. Они обеспечивают стабильность распределения входов в последующие слои. В GLU вход делится на две части: одна часть проходит через линейную функцию активации, а другая часть проходит через сигмоидную функцию активации, т. е. этот блок пропускает только половину информации из предыдущего слоя, взвешенную с помощью второй половины, прошедшей через сигмоиду, а затем происходит поэлементное умножение тензоров:

$$GLU(a, b) = a \odot \sigma(b),$$

где a – первая половина входного тензора; b – вторая; \odot – операция поэлементного умножения тензоров (разделение принято проводить по последней оси тензора); $\sigma(\cdot)$ – операция поэлементного применения сигмоиды.

В данной имплементации в каждом блоке GLU непосредственно самой операции предшествует линейный слой, увеличивающий количество признаков вдвое – это предотвращает затирание информации и сохранение размерностей. Сравнение представлено на рис. 2.

Блок смещения с нормализацией (Add & Norm) – это блок, который включает в себя слой нормализации и блок обхода (skip-connection). Блок смещения с нормализацией используется для комбинации текущего входа с предыдущим выходом и нормализации результирующего сигнала. Он представляет собой следующую последовательность действий:

- установка параметров блока: размерности входных и выходных тензоров, возможность обучения маски;
- если размерность входного тензора не равна размерности выходного тензора, то применяется блок интерполяции; в противном случае используется слой-тождественность;

- если параметр добавления обучаемый, то создается параметр маски, который затем проходит через блок сигмоидальной активации и после этого умножается на параметр пропуска и некоторый множитель;
- вычисление нормализации входного тензора с добавлением параметра пропуска;
- возвращение выходного тензора.

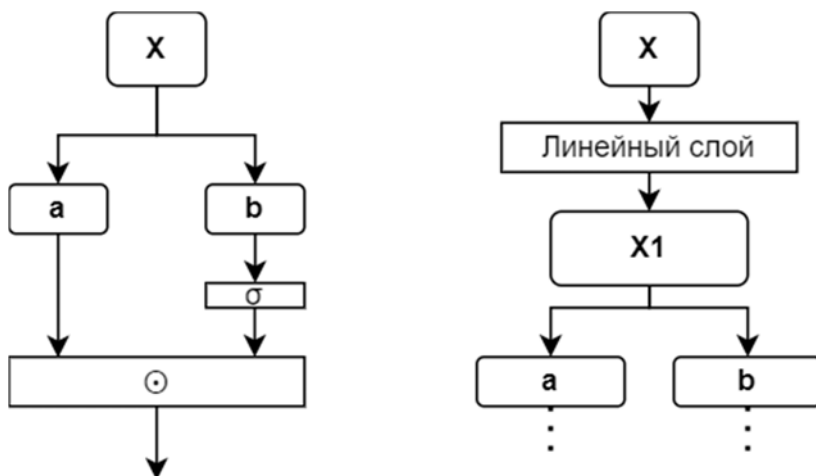


Рис. 2. Слева – исходный блок GLU, справа – используемый в данной имплементации, где разделению входов предшествует линейный слой, увеличивающий вход вдвое по последней оси тензора.

Fig. 2. Left plot - original GLU block, right plot - used in this implementation, where a linear layer is applied before the input split, doubling the input along the last axis of the tensor.

Слой Интерпретируемое Многоголовое Внимание (Interpretable MultiHead Attention) – это модификация стандартного механизма многоголового внимания [8], которая обеспечивает более прозрачную и интерпретируемую работу слоя. Этот блок принимает на вход три вектора: запросы (Q), ключи (K) и значения (V).

Каждый из этих векторов проходит через отдельные линейные преобразования, чтобы получить векторы с размерностью $\left(\frac{d_{model}}{heads}\right)$, где $heads$ – число голов, на которые разбивается вход этого слоя, d_{model} – размерность скрытого состояния модели. Затем происходит процесс «внимания», где каждая голова вычисляет веса «внимания» для каждой пары запросов и ключей, а затем объединяет взвешенные значения из вектора значений, чтобы получить итоговые значения для каждой головы. Далее, все головы объединяются и проходят через линейное преобразование, чтобы получить окончательный выход.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}} + mask\right)V$$

$$head_i, attn_i = Attention(Q, K, V, mask)$$

$$mask = M_{ij} = \begin{cases} 1, & \text{если } j \geq i + encLen, \\ i \in [0, decLen - 1], \quad j \in [0, encLen + decLen - 1]; \\ 0, & \text{в обратном случае;} \end{cases}$$

$$head = \frac{1}{heads} \sum_{i=0}^{heads-1} head_i, \quad attn = \frac{1}{heads} \sum_{i=0}^{heads-1} attn_i$$

$$preoutput = \frac{1}{heads} \sum_{i=0}^{heads-1} head_{:,i}, \quad \in \mathbb{R}^{B \times S \times d_v}$$

$$output = w_h(preoutput) \in \mathbb{R}^{B \times S \times d_{model}}$$

где d_{model} – размерность скрытого состояния модели; d_v – размерность вектора значений; d_k – размерность вектора ключей; B – размер пакета (batch size); S – размерность временного шага; i – индекс головы; $head$ – тензор выхода голов; $attn$ – тензор внимания; $mask$ – тензор бинарной маски; $encLen$ – длина энкодера; $decLen$ – длина декодера; w_h – слой линейного преобразования выхода голов; $preoutput$ – среднее арифметическое выходов голов; $output$ – выход слоя. В данной реализации используются отдельные линейные слои для запросов и ключей для каждой из голов, в то время как в стандартной имплементации все головы используют одинаковые линейные преобразования для Q , K и V . Кроме того здесь введен дополнительный слой w_h для объединения выходов голов, а также используется dropout для предотвращения переобучения на выходе данного слоя

$$output = m \odot x,$$

где x – входной тензор; m – бинарная маска, определяющая, какие элементы входного тензора следует обнулить; оператор \odot – поэлементное умножение. При обучении модели бинарная маска генерируется случайным образом с заданной конфигурацией модели вероятностью p для каждого входного тензора, в то время как во время предсказания $dropout$ не применяется.

Помимо архитектуры самой модели, достаточно внимания должно уделяться предобработке входов в модель. Ранее уже отмечалось, что нормализация может улучшить точность модели, поэтому для каждого непрерывного признака данных используется закрепленный за ним нормализатор (StandardScaler), обеспечивающий нормализованное распределение этого признака. Схематично описанный процесс представлен на рис. 3.

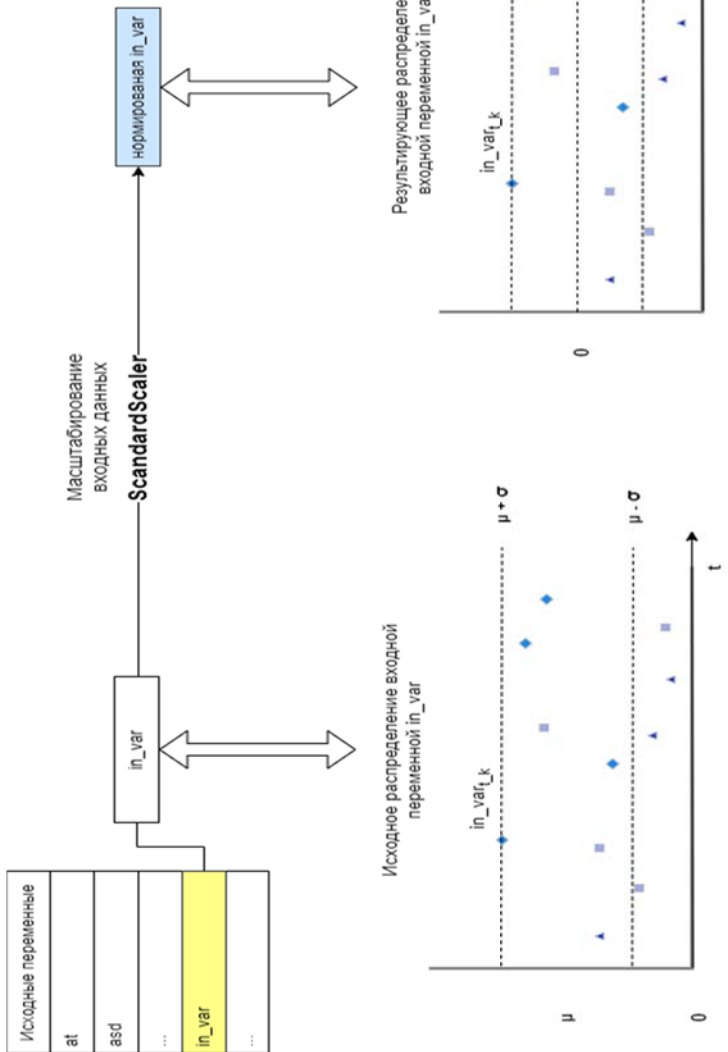


Рис. 3. Схема процесса нормализации входных данных (распределение переменной in_var со средним значением μ и стандартным отклонением σ переходит в стандартное распределение).
Fig. 3. Diagram of the input data normalization process (the distribution of the variable in_var with mean μ and standard deviation σ is transformed into a standard distribution).

В качестве дополнения для целевой переменной мы дополнительно удерживаем эти параметры нормализации (среднее и стандартное отклонение) и подаем их на вход в модель, где они используются для обратного масштабирования выходов модели (рис. 4).

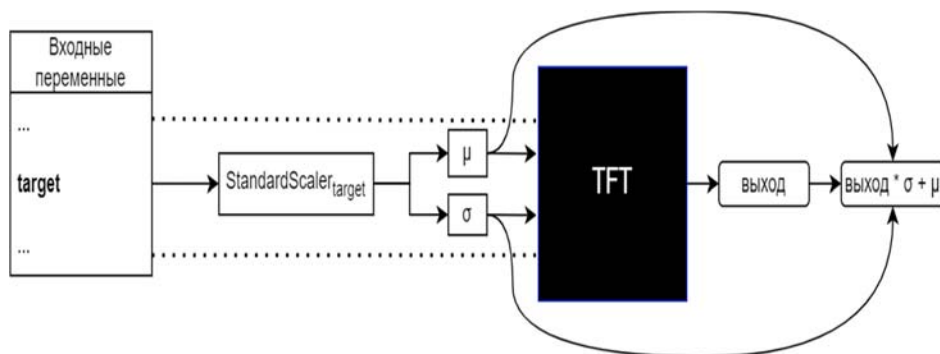


Рис. 4. Графическое представление процедуры преобразования целевой переменной target через соответствующий нормализатор.

Fig. 4. Visual representation of the target variable transformation using the corresponding normalizer.

2. Численная реализация и анализ результатов прогнозов

В качестве объекта численной реализации модифицированной модели TFT были выбраны три водпоста р. Исеть (бассейн речной системы р. Тобол): Катайск, Шадринск, Мехонское, для которых все классические методы долгосрочных прогнозов максимальных уровней воды оказались безуспешными. В общем виде задача долгосрочного прогноза максимального уровня воды в данном случае формулируется следующим образом. Имея временные ряды нескольких входных переменных, необходимо предсказать уровень воды на каждом временном шаге, интерпретируемом как 1 сутки. В табл. 1–3 представлено описание каждой из используемых переменных.

Все исходные данные для проведения численных экспериментов были представлены ФГБУ «Уральское УГМС» и ФГБУ «Гидрометцентр России». Использовалось следующее пространственное распределение гидрометеорологической информации:

1) гидрологические данные (среднесуточные расход и уровень воды) по водпостам Катайск, Шадринск, Мехонское;

2) метеорологическая информация (суточные температура воздуха, высота снежного покрова, скорость ветра, температура точки росы, сумма осадков, относительная влажность в %) по метеостанциям Далматово, Шадринск, Шатрово, Курган.

В зависимости от наличия пропусков и их количества, а также характера индивидуального поведения кривой какой-либо из переменных, для

заполнения данных использовались линейная, бикубическая интерполяция или же сплайны.

Таблица 1. Список исходных переменных, полученных из исходных данных, и их преобразование

Table 1. List of input features derived from the source data and their transformations

Кодировка	Физический смысл	Расчетная формула
at_N_h	Температура воздуха N часов, °C	-
at	Температура воздуха среднесуточная, °C	$\frac{1}{M} \sum_{\text{в течение суток}} at_N_h$
asd	Среднесуточная высота снежного покрова, см	-
ws_N_h	Средняя скорость ветра в срок наблюдения (каждые h часов), м/с	-
awspd	Среднесуточная скорость ветра, м/с	$\frac{1}{M} \sum_{\text{в течение суток}} ws_N_h,$
dpt_N_h	Температура точки росы в срок наблюдения, °C	-
dpt	Среднесуточная температура точки росы, °C	$\frac{1}{M} \sum_{\text{в течение суток}} dpt_N_h$
ps_N_h	Сумма осадков за период между сроками h часов, м	-
psbd	Сумма осадков за сутки, мм	$\frac{1}{M} \sum_{\text{в течение суток}} ps_N_h$
rhp_N_h	Относительная влажность в срок наблюдения (h часов), %	-
rhp	Среднесуточная влажность, от 0.0 до 1.0	$\frac{1}{M \cdot 100} \sum_{\text{в течение суток}} rel_hum_N_h$
flow	Среднесуточный расход воды, м ³ /сек	-
level	Среднесуточный уровень воды, в см над «0» графика поста	-

Примечание. M – количество показаний.

Таблица 2. Список дополнительно введенных переменных
Table 2. List of additional introduced features

Кодировка	Расчетная формула
month	Взятие месяца из текущей даты
month_x	$\sin(2\pi \cdot \text{month}/12)$
month_y	$\cos(2\pi \cdot \text{month}/12)$
day	Взятие дня из текущей даты с начала текущего года
day_x	$\sin(2\pi \cdot \text{day}/365) \mid \sin(2\pi \cdot \text{day}/366)$
day_y	$\cos(2\pi \cdot \text{day}/365) \mid \cos(2\pi \cdot \text{day}/366)$
relative_time_idx	$\frac{t}{\max_encoder_len'}$ <p>где t – позиция временного шага, $\max_encoder_len$ – максимальная длина энкодера (входная длина)</p>
level_scale	$\frac{\sigma(\text{level}) - \mu_{\text{level_scale}}}{\sigma_{\text{level_scale}}}$ <p>где $\sigma(\text{level})$ – взятие стандартного отклонения для текущего входного временного ряда level длиной encoder_length; $\mu_{\text{level_scale}}$ и $\sigma_{\text{level_scale}}$ – среднее и стандартное отклонение по всей выборке; последнее рассчитывается по совокупности стандартных отклонений посчитанных на группах длиной encoder_length + decoder_length</p>
level_center	$\frac{\mu(\text{center}) - \mu_{\text{level_center}}}{\sigma_{\text{level_center}}}$ <p>где $\mu(\text{level})$ – взятие среднего для текущего входного временного ряда level длиной encoder_length; $\mu_{\text{level_center}}$ и $\sigma_{\text{level_center}}$ – среднее и стандартное отклонение среднего по всей выборке; последнее рассчитывается по совокупности средних посчитанных на группах длиной encoder_length + decoder_length.</p>
encoder_length	Длина входного временного ряда
region	Закодированное (в целых числах) имя водпоста

Каждый из гидрологических постов ассоциировался с данными одной или нескольких метеостанций и группировка по дате с усреднением характеристик. В наших расчетах использовалось соответствие, приведенное в табл. 4.

Все модели обучались на зависимой выборке с 1991 по 2017 г. (27 лет) и валидировались на независимой выборке 2018–2022 гг. (5 лет) с одинаковыми параметрами обучения, представленными в табл. 5.

Таблица 3. Типы переменных, подаваемых на вход модифицированной модели**Table 3.** Feature types list for the modified model

Кодировка переменной	Тип переменной
encoder_length	Статическая непрерывная
level_center	Статическая непрерывная
level_scale	Статическая непрерывная
time_idx	Временная непрерывная, известная заранее
relative_time_idx	Временная непрерывная, известная заранее
at	Временная непрерывная, наблюдаемая
asd	Временная непрерывная, наблюдаемая
awspd	Временная непрерывная, наблюдаемая
dpt	Временная непрерывная, наблюдаемая
psbd	Временная непрерывная, наблюдаемая
rhp	Временная непрерывная, наблюдаемая
level	Временная непрерывная, наблюдаемая
flow	Временная непрерывная, наблюдаемая
month	Временная категориальная, известная заранее
month_x	Временная непрерывная, известная заранее
month_y	Временная непрерывная, известная заранее
day_x	Временная непрерывная, известная заранее
day_y	Временная непрерывная, известная заранее
region	Статическая категориальная

Таблица 4. Соответствие данных наблюдений по исходным водпостам и ассоциированным метеостанциям**Table 4.** Mapping of observation data between stream gauging stations and associated weather stations

Водпост	Ассоциированные метеостанции
Катайск	Далматово
Шадринск	Шадринск
Мехонское	Шатрово, Курган, Шадринск

В качестве оптимизатора был использован Adam (Adaptive moment estimation) с параметрами $\epsilon = 10^{-8}$, $\beta_0 = 0,9$, $\beta_1 = 0,999$, а в качестве функции потерь Quantile Loss [7].

Таблица 5. Общие гиперпараметры для обучения моделей
Table 5. General model hyperparameters

Название гиперпараметра	Описание	Значение
batch_size	Размер пакета при обучении	256
max_epochs	Максимальное количество эпох при обучении	20
learning_rate	Скорость обучения	1e-4
reduce_lr_every_n_epochs	Частота снижения скорости обучения	4
lstm_layers	Количество слоев LSTM	2
lr_scheduler_gamma	Коэффициент затухания скорости обучения	0.15
hidden_size	Размер скрытого слоя	256
hidden_continuous_size	Размер скрытого слоя для непрерывных признаков	256
attention_head_size	Число голов внимания	8
Params	Количество обучаемых параметров, млн	8,9

Дополнительно по запросу ФГБУ «Уральское УГМС» в 2023 г. был выпущен оперативный прогноз максимальных уровней воды с использованием оперативных данных гидрометеорологических наблюдений в бассейне р. Исеть. Выпуск оперативного прогноза проводился с теми же гиперпараметрами, что и для других лет наблюдений.

Реализация модели для всех указанных расчетных лет проводилась для двух вариантов конфигурации обучения модели с суточным интервалом по времени:

- первый вариант (Var. I) – это 90 суточных значений до даты прогноза на входе и 60 суточных значений после даты прогноза на выходе;
- второй вариант (Var. II) – 90 значений на входе и 90 значений на выходе.

Оправдываемость долгосрочных прогнозов максимальных уровней воды оценивалась по следующим критериям:

• $\Delta_i^j = \frac{abs(y_i^j - \widehat{y}_i^j)}{\Delta_{доп}}$ – относительная ошибка прогноза максимального уровня воды при заданной его допустимой погрешности для j -го водпоста и i -го года, где y_i^j – фактическое значение уровня; \widehat{y}_i^j – его предсказанное значение; $\Delta_{доп}$ – допустимое значение погрешности прогноза максимального уровня для каждого j -го водпоста, заданное ФГБУ «Уральское УГМС». Оправданность прогноза максимального уровня определяется принадлежностью оценки Δ_i^j по каждому водпосту для одного из четырех интервалов ($K_i^j(s)$): $[0, 0.4)$ – отлично ($s = 1$); $[0.4, 0.7)$ – хорошо ($s = 2$); $[0.7, 1.0)$ – удовлетворительно ($s = 3$); $[1.0, +\infty)$ – неудовлетворительно ($s = 4$);

• $\Delta_i^t(j) = abs(T_{mi} - T_{mfi})$ – абсолютная ошибка даты наступления максимального уровня воды для каждого j -го водпоста по каждому прогнозируемому i -му году (в сутках), где $T_{mi}(j)$, $T_{mfi}(j)$, – соответственно, даты наблюдаемого и прогнозируемого максимумов на j -м водпосту для i -го года.

В табл. 6 приведены результаты количественной оценки долгосрочных прогнозов максимального уровня воды для трех водпостов р. Исеть по двум вариантам конфигурации модели (Var. I, Var. II) при следующих допустимых значений погрешности прогноза, полученных из ФГБУ «Уральское УГМС»: 1) Катайск – 80 см; 2) Шадринск – 69 см; 3) Мехонское – 56 см. Графическое сопоставление фактических и предсказанных среднесуточных значений уровней воды по трем водпостам представлены на рис. 5–7.

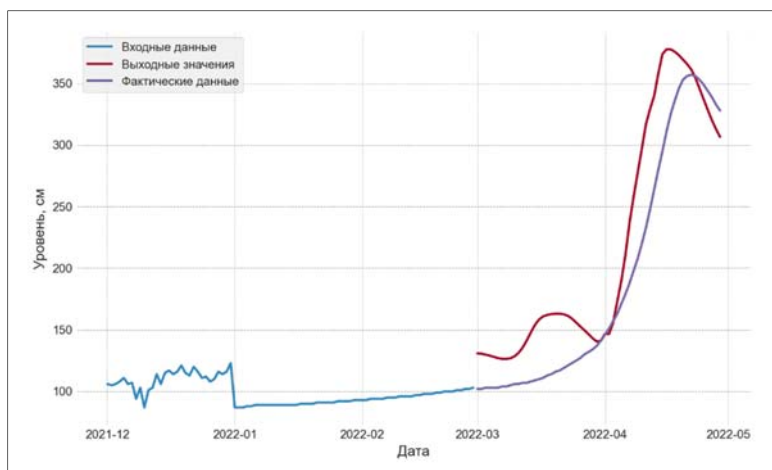


Рис. 5. Фактические и предсказанные значения уровней воды для водпоста Катайск (2022 г.)

Fig. 5. Actual and predicted water level values for the Kataysk stream gauging station (2022 year).

Таблица 6. Значения количественных критериев оценки качества выпущенных долгосрочных прогнозов максимальных уровней воды р. Исеть с использованием модифицированной модели TFT
Table 6. Quantitative evaluation criteria values for the long-term forecasts of maximum water levels of the Iset River using the modified TFT mode

Название водпоста	Расчетный прогнозный год	Δ_t^j		Качественная оценка прогнозов по интервальной принадлежности их отн. ошибки ($K_t^j(s)$)	Число прогнозов, которые оправдались по данным для независимой выборки, в %		$\Delta_t^j(j)$, в сут		Средн. арифм. абсолют. ошибка прогнозной даты наступления макс. уровня воды, в сут		
		Var. I	Var. II		Var. I	Var. II	Var. I	Var. II	Var. I	Var. II	
Оценки для независимой выборки (2018-2022 гг.)											
Катайск	2018	1,1	2,4	4	4	60%	20%	11	12	9	10
	2019	0,5	2,4	2	4			5	9		
	2020	0,3	3,4	1	4			12	17		
	2021	1,5	0,7	4	3			6	7		
	2022	0,6	1,9	2	4			9	6		
Шадринск	2018	0,6	2,0	2	4	60%	40%	2	1	15	15
	2019	1,2	0,1	4	1			1	3		
	2020	0,7	2,5	3	4			16	18		
	2021	1,4	0,3	4	1			5	5		
	2022	0,2	2,5	1	4			50	49		
Мехонское	2018	0,5	0,9	2	3	60%	80%	3	2	10	9
	2019	2,8	0,7	4	3			3	1		
	2020	0,1	1,4	1	4			16	19		
	2021	2,1	0,5	4	2			4	2		
	2022	0,2	0,1	1	1			24	21		
Оценки оперативных прогнозов по данным за 2023 г.											
Катайск	2023	0,7	2,1	3	4	67%	67%	17	20	19	20
Шадринск	2023	1,7	0,6	4	2	67%	67%	18	22	19	20
Мехонское	2023	0,3	0,1	1	1			21	18		

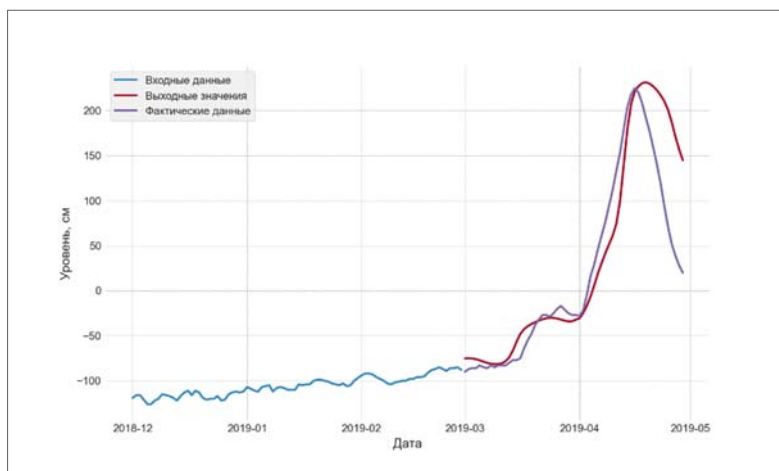


Рис. 6. Фактические и предсказанные значения уровней воды для водпоста Шадринск (2019 г.)

Fig. 6. Actual and predicted water level values for the Shadrinsk stream gauging station (2019 year).

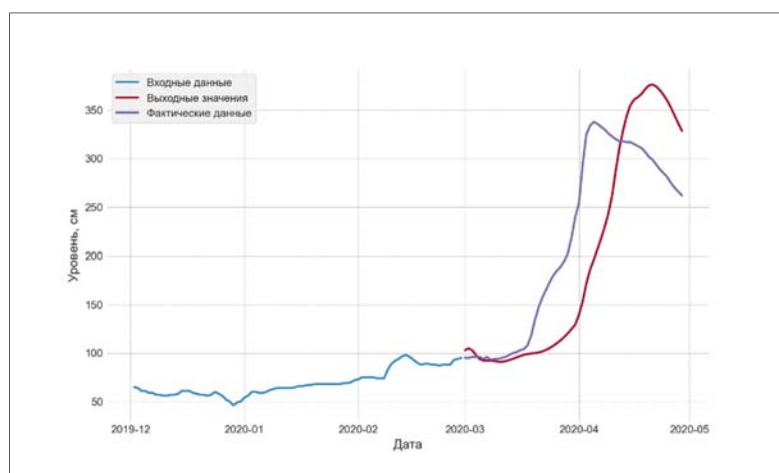


Рис. 7 Фактические и предсказанные значения уровней воды для водпоста Мехонское (2020 г.)

Fig. 7. Actual and predicted water level values for the Mekhonskoye stream gauging station (2020 year).

Заключение

Несмотря на пока еще крайне ограниченный объем выполненных численных экспериментов уже сейчас можно с уверенностью сказать, что использование ANN в ближайшем будущем позволит существенно пересмотреть наши представления о возможностях методов гидрологических

прогнозов водного режим. Неслучайно эта область выделена как одна из ведущих в рамках реформирования оперативной гидрологии, входящей в группу системных и структурных изменений ВМО ООН.

Целесообразно в общем виде выделить несколько комплексных перспективных направлений, связанных с реализацией алгоритмов глубокого машинного обучения для развития гидрологических прогнозов:

1. Расширение в использовании новой гидрологической информации при обучении исходной модели. В частности, примером такой информации могут служить данные измерения уровней грунтовых вод, которые измеряются на стационарных скважинах системы Роснедра, хорошо отражают увлажненность водосбора, реально доступны при решении некоторых организационных процедур и в силу своей комплексности могут заменить ряд гидрометеорологических характеристик, использованных при ее обучении.

2. Постепенный переход к использованию модели TFT для краткосрочных прогнозов водного режима с значительным расширением ее пространственного использования по территории Российской Федерации, а также в задачах долгосрочного прогноза минимального уровня воды в период межени.

3. Существенно более глубокий численный анализ возможностей используемой модели за счет: а) введение статистических переменных на период заблаговременности прогноза в виде месячной нормы осадков / прогнозной месячной величины осадков; б) независимое обучение модели по каждому водопосту; в) использование для обучения модели данных по максимальным уровням воды, фиксируемым сотрудниками УГМС, которые в ряде случаев существенно отличаются от среднесуточных значений, публикуемых в гидрологических ежегодниках; г) привлечение в качестве входных данных результатов численных прогнозов метеоэлементов по моделям циркуляции атмосферы, а также спутниковых снимков территории водосбора.

Можно предположить, что уже завтра указанный перечень окажется далеко неполным, учитывая стремительность обновления информации, связанной с развитием ANN.

Список литературы

1. Романов А.В. Дорожная карта современных гидрологических прогнозов водного режима // Метеорология и гидрология. – 2023 (в печати).
2. *Bhanja S., Das A.* Impact of data normalization on deep neural network for time series forecasting // arXiv preprint arXiv:1812.05519, 2018.
3. *Caldas F.M., Soares C.A* Temporal Fusion Transformer for Long-term Explainable Prediction of Emergency Department Overcrowding // Machine Learning and Principles and Practice of Knowledge Discovery in Databases: International Workshops of ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part I. – Cham: Springer Nature Switzerland, 2023. С. 71-88.
4. *Dauphin Y.N. et al.* Language modeling with gated convolutional networks // International conference on machine learning. PMLR, 2017. С. 933-941.

5. Hochreiter S., Schmidhuber J. Long short-term memory // *Neural computation*. 1997. Vol. 9, no. 8. P. 1735-1780.
6. jdb78, Temporal Fusion Transformer, [Online]. Available: https://github.com/jdb78/pytorch-forecasting/tree/master/pytorch_forecasting/models/temporal_fusion_transformer. [Accessed: Feb. 16, 2023].
7. Lim B. et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting // *International Journal of Forecasting*. 2021. Vol. 37, no. 4. P. 1748-1764.
8. Vaswani A. et al. Attention is all you need // *Advances in neural information processing systems*. 2017. T. 30.

References

1. Romanov A.V. Dorozhnaya karta sovremennyh gidrologicheskikh prognozov vodnogo rezhima. *Meteorologiya i Gidrologiya* [Russ. Meteorol. Hydrol.], 2023 – in print [in Russ.].
2. Bhanja S., Das A. Impact of data normalization on deep neural network for time series forecasting. arXiv preprint arXiv: 1812.05519, 2018.
3. Caldas F.M., Soares C.A Temporal Fusion Transformer for Long-term Explainable Prediction of Emergency Department Overcrowding. *Machine Learning and Principles and Practice of Knowledge Discovery in Databases: International Workshops of ECML PKDD 2022*, Grenoble, France, September 19–23, 2022, Proceedings, Part I. – Cham: Springer Nature Switzerland, 2023, pp. 71-88.
4. Dauphin Y.N. et al. Language modeling with gated convolutional networks // *International conference on machine learning*. PMLR, 2017, pp. 933-941.
5. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural computation*, 1997, vol. 9, no. 8, pp. 1735-1780.
6. jdb78, Temporal Fusion Transformer, [Online]. Available: https://github.com/jdb78/pytorch-forecasting/tree/master/pytorch_forecasting/models/temporal_fusion_transformer. [Accessed: Feb. 16, 2023].
7. Lim B. et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 2021, vol. 37, no. 4, pp. 1748-1764.
8. Vaswani A. Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need. *Advances in neural information processing systems*, 2017, vol. 30.

Поступила 15.05.2023; одобрена после рецензирования 30.05.2023;
принята в печать 13.06.2023.

Submitted 15.05.2023; approved after reviewing 30.05.2023;
accepted for publication 13.06.2023.