

Анализ LSTM и GRU моделей для построения прогнозов временных рядов

С. В. Козлов, С. А. Седенков

Аннотация – В статье представлен анализ применения рекуррентных нейронных сетей как инструмента прогнозирования временных рядов. Кратко изложены общие принципы работы рекуррентных нейронных сетей. Описаны преимущества их использования по сравнению со стандартными нейронными сетями и сверточными сетями. Рассмотрены области приложения разных видов архитектуры рекуррентных нейронных сетей. Проанализирован алгоритм функционирования рекуррентных нейронных сетей. Определен основной класс в описании алгоритма, приведена программная реализация его функций. Особое внимание уделено матричной форме параметров при выполнении алгоритма. Основная часть работы отведена сравнительному анализу модели длинной цепи элементов краткосрочной памяти и модели управляемых рекуррентных блоков. В статье кратко описана история их разработки, принцип работы каждой модели. Рассмотрены возможности этих моделей при решении задач построения временных рядов. Раскрыта сущность формул, с помощью которых нейронные сети LSTM и GRU производят вычисления. Охарактеризован программный код, разработанный авторами для каждой из моделей. Для анализа работы алгоритмов выбрано прогнозирование курса акций IBM. Данные, полученные в ходе эксперимента, приведены на графиках. В завершение работы дан их сравнительный анализ. Актуальность статьи обусловлена эффективностью внедрения методов рекуррентного анализа данных временных рядов с помощью нейронных сетей.

Ключевые слова – рекуррентная нейронная сеть, длинная цепь элементов краткосрочной памяти, управляемые рекуррентные блоки, моделирование, прогнозирование стоимости котировок, программа, нейросетевые технологии, информационные технологии.

I. ВВЕДЕНИЕ

В условиях текущей экономической ситуации и стремительного ускорения темпов научно-

Статья получена 16 мая 2024.

Козлов Сергей Валерьевич, Смоленский государственный университет, доцент кафедры прикладной математики и информатики, кандидат педагогических наук, доцент (email: svkozlov1981@yandex.ru)

Седенков Сергей Александрович, Смоленский государственный университет, студент физико-математического факультета (email: ejik200014@gmail.com)

технического развития вопросы планирования и принятия решений на основе прогнозирования [1, 2] приобретают все большее значение. Прогностика как наука активно развивается в последние десять лет, что служит результатом создания множества методов [3, 4], процедур, приемов прогнозирования, абсолютно различных по своему значению и способу применения. Сейчас насчитывается свыше ста методов построения прогноза. В связи с этим, перед экспертами возникает вопрос актуальности выбора метода, который смог бы дать наиболее качественный прогноз для изучаемых процессов или систем.

Задача прогнозирования временных рядов является наиболее актуальной экономической задачей, так как у человека, желающего получить прибыль на бирже, возникает потребность в изучении данных о состоянии торговли в прошлом с целью оценки будущих результатов работы. До недавнего времени большинство людей использовали статистические методы для построения прогнозов временных рядов. Но математические модели не всегда могли с точной вероятностью спрогнозировать сложные явления и процессы [5, 6], в которых модель данных может быть нелинейной. Именно из-за таких случаев эксперты стали чаще прибегать к использованию аппарата нейронных сетей.

Нейронные сети являются математическими инструментами, позволяющими моделировать разного рода зависимости, такие как линейные модели [7, 8], обобщенно линейные модели, нелинейные модели [9, 10]. В настоящее время одним из основных востребованных направлений использования нейронных сетей является моделирование нелинейных зависимостей [11, 12]. Нейронная сеть способна к обобщению и выделению скрытых зависимостей между входными и выходными данными, что приводит к такой ее функциональной способности как прогнозирование. Нейронная сеть, обученная корректно, способна предсказать будущие значения исследуемых в настоящий момент времени факторов на основе их предыдущих значений.

В данной статье описаны основные принципы работы современных рекуррентных нейронных сетей (RNN). Более подробно описаны такие модели как: длинная цепь элементов краткосрочной памяти (LSTM) и управляемые рекуррентные блоки (GRU). Разработан и представлен программный комплекс для прогнозирования стоимости котировок. Также приводится сравнение двух моделей (LSTM и GRU) между собой.

II. РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ

Рекуррентные нейронные сети (Recurrent Neural Networks, RNN) – нейронные сети, в которых связи между элементами имеют направленную последовательность [13]. Из-за этого возникает возможность обрабатывать целые ряды событий во времени [14]. Рекуррентные нейронные сети используют свою внутреннюю память для обработки последовательностей произвольной длины, поэтому они часто используются в таких задачах, где нечто целостное разбито на части, например, в задачах распознавания изображений [15, 16] или распознавания речи [17, 18].

Одной из основных проблем стандартных нейронных сетей (Neural Networks, NN) и сверточных сетей (Convolutional Neural Network, CNN) [19] является чрезмерная ограниченность: в роли входных данных выступает вектор фиксированного размера, например, изображение или текст. Далее нейронная сеть воспроизводит вектор фиксированного размера в качестве вывода, например, вероятности различных классов. Кроме того, модели выполняют сопоставление, используя фиксированное количество вычислительных шагов, например, количество слоев в модели. RNN позволяют работать над последовательностями векторов на входе и выходе или над тем и другим. На рисунке 1 представлены различные вариации, структуры RNN.

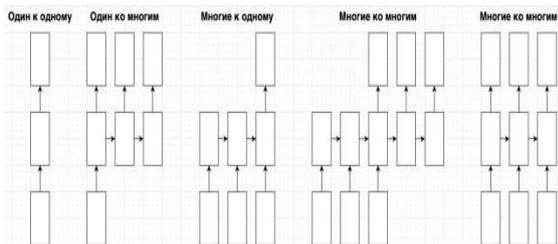


Рис. 1. Вариации архитектуры сетей RNN

Каждый прямоугольник представлен в виде вектора, а стрелки являются функциями, например, умножением или сложением матриц [20]. Слева направо. Структура «один к одному» представляет входные и выходные данные фиксированного размера, например, классификация изображений [21]. Структура «один ко многим» – выход последовательности, например, субтитры к изображению; в роли входных данных поступает изображение, а на выходе получается набор слов. Структура «многие к одному» задает последовательность ввода, например, анализ настроений [22], когда на ввод поступает предложение, которое должно получить на выходе готовый эмоциональный окрас. При этом выражение классифицируется как набор положительных или отрицательных настроений. Первая из структур «многие ко многим» определяет входные данные и последовательности, например, машинный перевод. В роли входных данных поступает предложение на русском языке, а затем происходит перевод предложения на английский язык [23]. Вторая структура «многие ко многим» обрабатывает синхронизированные входные и выходные данные

последовательности, например, классификация видео, где можно пометить каждый кадр видео [24].

Алгоритм RNN довольно простой: в роли входных данных подается вектор X , а на выходе получаем вектор Y . На содержимое выходного вектора влияют не только входные параметры, но и все исторические данные входных параметров, которые использовались в прошлом. Класс алгоритма RNN представлен одношаговой функцией:

```
rnn = RNN()
y = rnn.step(x) # x – входной вектор, y – выходной вектор RNN
```

Класс RNN имеет некоторое внутреннее состояние, которое обновляется каждый раз, когда вызывается функция $step()$. В простейшем случае функция состояния состоит из скрытого вектора h . Реализация функции $step()$ в стандартном RNN алгоритме имеет вид:

```
class RNN:
    # ...
    def step(self, x):
        # обновляем скрытое состояние
        self.h = np.tanh(np.dot(self.W_hh, self.h) +
            np.dot(self.W_xh, x))
        # вычисляем выходной вектор
        y = np.dot(self.W_hy, self.h) return y
```

В роли параметров RNN выступают матрицы: W_{hh} , W_{xh} , W_{hy} . Скрытое состояние $self.h$ обозначается нулевым вектором. Функция $np.tanh$ демонстрирует нелинейность, которая ограничивает возможные значения в диапазоне $[-1, 1]$. Функция $tanh$ имеет два параметра первый зависит от скрытого состояния, а второй от входного параметра.

Матрица RNN представлена случайными числами, и большая часть работы во время обучения уходит на поиск матриц, которые приводят к желаемому поведению, измеренному с помощью некоторой функции потерь, которая выражает предпочтения выходных типов данных y на входные последовательности x .

Способность обрабатывать последовательности делает RNN очень полезным в задачах машинного перевода, например, Google Translate или Yandex Translate. В таком случае текстовая последовательность подается в качестве входных данных и преобразуется в текст на нужном языке. RNN также часто применяют для анализа эмоционального окраса, например, выявления положительного или негативного отзыва [25]. Текст подается на вход, затем производится выходная кластеризация [26].

III. МОДЕЛЬ ДЛИННОЙ ЦЕПИ ЭЛЕМЕНТОВ КРАТКОСРОЧНОЙ ПАМЯТИ (LSTM-МОДЕЛЬ)

Наиболее мощная архитектура RNN – это длинная цепь элементов краткосрочной памяти (Long short-term memory; LSTM), разработана в 1997 году З. Хохрайтером и Ю. Шмидхубером. Также, как и большинство рекуррентных нейронных сетей, LSTM-

сеть универсальна, в том смысле, что достаточное количество сетевых элементов позволяет выполнять все вычисления, которые может выполнить обычный компьютер. В отличие от традиционных рекуррентных нейронных сетей, LSTM-модель отлично адаптирована к обучению на задачах классификации, обработки и прогнозированию временных рядов в случаях, когда важные события разделены временными разрывами с неопределённой продолжительностью и границами. LSTM-модель имеет весомое преимущество по отношению к альтернативным рекуррентным нейронным сетям, и другим методам обучения для последовательностей, из-за невосприимчивости к временным задержкам и разрывам. Модель легко обучается и применяется в различных сферах деятельности, например в сфере финансов [27]. В качестве примера можно выделить одно из главных достижений LSTM-сети. Это наилучшие результаты в распознавании несегментированного слитного рукописного текста и победу в 2009 году на соревнованиях по распознаванию рукописного текста (ICDAR). LSTM-сети также находят широкое применение в использовании в задачах распознавания речи. В качестве примера можно рассмотреть случай, когда LSTM-сеть была основным компонентом сети, которая в 2013 году достигла рекордного порога ошибки в 17,7% в задаче распознавания фонов на классическом корпусе естественной речи TIMIT [28]. На рисунке 2 представлена графическая архитектура устройства LSTM-сети.

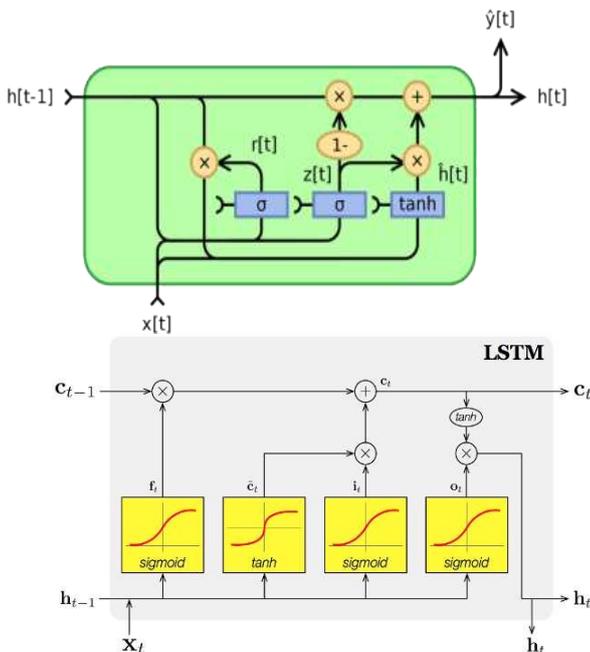


Рис. 2. Схематичное изображение архитектуры LSTM-сети

Описание LSTM-модели с помощью формул имеет следующий вид:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$\begin{aligned} i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i), \\ o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o), \\ c_t &= f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c), \\ h_t &= o_t \circ \sigma_h(c_t), \end{aligned}$$

Пояснение обозначений:

$c_0 = 0$ и $h_0 = 0$ – векторы вентиля забывания;

\circ – произведение Адамара;

x_t – входной вектор;

h_t – выходной вектор;

c_t – вектор состояний; W , U и b – матрицы параметров и вектор;

f_t – вектор вентиля забывания, вес запоминания старой информации;

i_t – вектор входного вентиля, вес получения новой информации;

o_t – вектор выходного вентиля, кандидат на выход;

σ_g – функция активации на основе сигмоиды;

σ_c – функция активации на основе гиперболического тангенса;

σ_h – функция активации на основе гиперболического тангенса.

IV. УПРАВЛЯЕМЫЕ РЕКУРРЕНТНЫЕ БЛОКИ (GRU)

Первый вентиляльный механизм для рекуррентных нейронных сетей был представлен в 2014 году. При решении задачи моделирования музыкальных и речевых сигналов [29] было обнаружено, что их эффективность сопоставима с использованием долгой краткосрочной памяти (LSTM). Из-за отсутствия выходного вентиля данный механизм имеет меньше параметров чем LSTM-модель (рис. 3). У блока GRU нет необходимости использования блока памяти для управления потоком информации, что является существенным отличием по сравнению с LSTM-моделью. В блоке GRU используются напрямую все скрытые состояния без какого-либо контроля. Так как GRU имеет меньше параметров, то модель может обучаться быстрее. При этом она может нуждаться в меньшем количестве данных для обобщения. Но с большим количеством информации эксперты зачастую прибегают к использованию LSTM-модели [30].

Рис. 3. Схематичное изображение архитектуры LSTM-сети

Описание GRU-модели с помощью формул имеет следующий вид:

$$\begin{aligned} z_t &= \sigma_g(W_z x_t + U_z h_{t-1} + b_z), \\ r_t &= \sigma_g(W_r x_t + U_r h_{t-1} + b_r), \\ h_t &= z_t \circ h_{t-1} + (1 - z_t) \circ (W_h x_t + U_h (r_t \circ h_{t-1}) + b_h), \end{aligned}$$

Пояснение обозначений:

x_t – входной вектор; h_t – выходной вектор;

z_t – вектор вентиля обновления;

r_t – вектор вентиля сброса; W , U и b – матрицы параметров и вектор,
 σ_g – функция активации на основе сигмоиды;
 σ_h – функция активации на основе гиперболического тангенса.

V. ПОСТРОЕНИЕ ПРОГНОЗА СТОИМОСТИ КОТИРОВОК ПРИ ПОМОЩИ LSTM И GRU СЕТЕЙ

В ячейке LSTM-модели присваивается предыдущее состояние памяти c_{t-1} и выполняется поэлементное умножение с вентиляем забывания f_t , чтобы определить, присутствует ли состояние памяти c_t . Если значение вентиля забывания равно 0, то предыдущее состояние памяти полностью стирается, иначе значение вентиля забывания равно 1, тогда предыдущее состояние памяти полностью передается в ячейку:

$$c_t = c_{t-1} \cdot f_t$$

$$c_t = c_t + (i_t \cdot c_t')$$

$$H_t = \tanh c_t$$

Приведем пример кода для LSTM-модели:

```
# Архитектура LSTM
regressor = Sequential()
regressor.add(LSTM(units=50,
return_sequences=True,
input_shape=(X_train.shape[1],1)))
regressor.add(Dropout(0.2))
regressor.add(LSTM(units=50,
return_sequences=True))
regressor.add(Dropout(0.2))
regressor.add(LSTM(units=50,
return_sequences=True))
regressor.add(Dropout(0.2))
regressor.add(LSTM(units=50))
regressor.add(Dropout(0.2))
regressor.add(Dense(units=1))
# Компилирование RNN
regressor.compile(optimizer='rmsprop',loss='mean_squared_error')
regressor.fit(X_train,y_train,epochs=50,batch_size=32)
```

Подготовка тестового набора аналогично обучающему – для 60 целых наборов есть 60 предыдущих значений, которые невозможно получить без остальных, параметр – High

```
dataset_total = pd.concat((dataset["High"][:'2021'],dataset["High"]['2022':]),axis=0)
inputs = dataset_total[len(dataset_total)-len(test_set) - 60:].values
inputs = inputs.reshape(-1,1)
inputs = sc.transform(inputs)
```

```
# Подготовка и прогнозирование
X_test = []
for i in range(60,311): X_test.append(inputs[i-60:i,0])
X_test = np.array(X_test)
```

```
X_test = np.reshape(X_test,
(X_test.shape[0],X_test.shape[1],1))
predicted_stock_price = regressor.predict(X_test)
predicted_stock_price = sc.inverse_transform(predicted_stock_price)
```

```
# Визуализация результатов
plot_predictions(test_set,predicted_stock_price)
# Оценка модели
return_rmse(test_set,predicted_stock_price)
```

Результат полученного прогноза при помощи LSTM-модели продемонстрирован на рисунке 4.

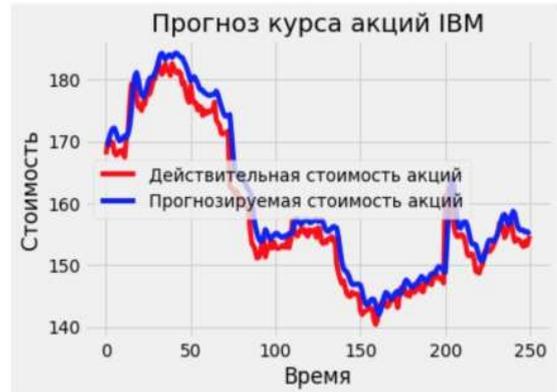


Рис. 4. Результат выполнения прогноза с помощью LSTM-модели

GRU в общих чертах похож на LSTM, исключение составляет то, что у GRU задействуются два шлюза: шлюз сброса и шлюз обновления. Шлюз сброса определяет, как объединить новый ввод с предыдущей памятью, а шлюз обновления определяет, какую часть предыдущего состояния следует сохранить.

Приведем пример кода для GRU-модели:

```
# Архитектура GRU
regressorGRU = Sequential()
# Первый слой GRU
regressorGRU.add(GRU(units=50,
return_sequences=True,
input_shape=(X_train.shape[1],1), activation='tanh'))
regressorGRU.add(Dropout(0.2))
```

```
# Второй слой GRU
regressorGRU.add(GRU(units=50,
return_sequences=True,
input_shape=(X_train.shape[1],1), activation='tanh'))
regressorGRU.add(Dropout(0.2))
```

```
# Третий слой GRU
regressorGRU.add(GRU(units=50,
return_sequences=True,
input_shape=(X_train.shape[1],1), activation='tanh'))
regressorGRU.add(GRU(units=50,
activation='tanh'))
regressorGRU.add(Dropout(0.2))
# Выходной слой
regressorGRU.add(Dense(units=1))
```

```
# Создание RNN сети
```

```
regressorGRU.compile(optimizer=SGD(lr=0.01,
decay=1e-7, momentum=0.9,
nesterov=False),loss='mean_squared_error')
```

```
# Подготовка X_test и прогнозирование цен
X_test = []
for i in range(60,311): X_test.append(inputs[i-
60:i,0])
X_test = np.array(X_test)
X_test = np.reshape(X_test,
(X_test.shape[0],X_test.shape[1],1))
GRU_predicted_stock_price = regressorGRU.predict(X_test)
GRU_predicted_stock_price = sc.inverse_transform(GRU_predicted_stock_price)

# Визуализация результатов
plot_predictions(test_set,GRU_predicted_stock_price)

# Оценка модели
return_rmse(test_set,GRU_predicted_stock_price)
```

Результат полученного прогноза при помощи GRU-модели продемонстрирован на рисунке 5.

Рис. 5. Результат выполнения прогноза с помощью GRU-модели

ЗАКЛЮЧЕНИЕ

В ходе работы был проведен анализ и выявлены различия между RNN модулями: LSTM и GRU. Было выяснено, что GRU-модель использует меньше параметров при обучении и использует меньше памяти. Сам процесс обучения модели выполняется быстрее, чем LSTM, в то время как LSTM более точна для большего объема данных. Как правило, LSTM используется, если есть необходимость работы с большими последовательностями и необходим точный расчет, в то время как GRU используется, при необходимости получения более быстрого результата обучения с наименьшим потреблением памяти.

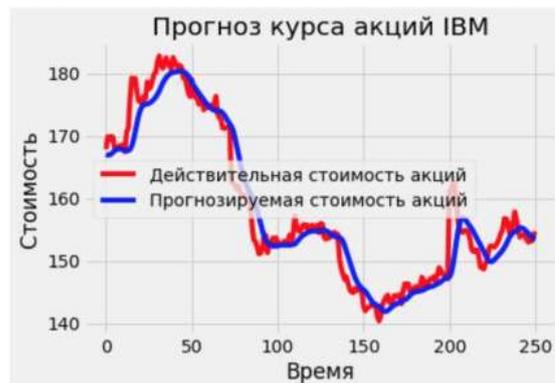
По полученным в ходе исследования результатам расчетов среднеквадратичная ошибка LSTM составляет не более 4-5%, в то время как среднеквадратичная ошибка GRU – не более 9% по сравнению с абсолютными значениями. Из этого следует вывод, что модели можно использовать для получения высокоточных результатов, а также применять для расчетов реального прогнозирования.

БИБЛИОГРАФИЯ

- [1] Антонов В. В., Пальчевский Е. В., Еникеев Р. Р. Прогнозирование на основе искусственной нейронной сети второго поколения для поддержки принятия решений в особо значимых ситуациях // Программные продукты и системы. – 2022. – № 3. – С. 384-395.
- [2] Козлов С. В. Использование функциональных возможностей информационных систем в производственной сфере // ЭНЕРГЕТИКА,

ИНФОРМАТИКА, ИННОВАЦИИ – 2017 (электроэнергетика, электротехника и теплоэнергетика, математическое моделирование и информационные технологии в производстве). Сборник трудов VII-ой Международной научно-технической конференции. – 2017. – В 3 т. Т 1. – С. 298-301.

- [3] Фахрутдинова А. З. Методологические проблемы современной прогностики // Омские научные чтения. Материалы Всероссийской научно-практической конференции. – 2017. – С. 851-853.
- [4] Козлов С. В., Суин И. А. О некоторых аспектах применения инвариантных методов функционального анализа данных в различных предметных областях // Системы компьютерной математики и их приложения. – 2019. – № 20-1. – С. 199-205.
- [5] Дубенко Ю. В., Дышкант Е. Е. Нейросетевой алгоритм выбора методов для прогнозирования временных рядов // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2019. № 1. С. 51-60.
- [6] Козлов С. В. Использование соответствия Галуа как инварианта отбора контента при проектировании информационных систем // Современные



информационные технологии и ИТ-образование. – 2015. – Т. 2. № 11. – С. 220-225.

- [7] Алымова Е. В. Совместное применение модели линейной регрессии и нейронной сети в задаче предсказания тренда котировок криптовалюты Bitcoin // Инженерный вестник Дона. – 2020. – № 10 (70). – С. 90-96.
- [8] Сенотова С. А. Сравнительный анализ методов аппроксимации с помощью регрессионных зависимостей и нейронных сетей для линейных моделей // Сборник научных трудов Ангарского государственного технического университета. – 2021. – Т. 1. № 18. – С. 31-35.
- [9] Соловьева Е. Б. Рекуррентные нейронные сети в качестве моделей нелинейных динамических систем // Цифровая обработка сигналов. – 2018. – № 1. – С. 18-27.
- [10] Дуболазов В. А., Сомов А. Г. Создание нелинейных аппроксимаций современных экономических моделей методом нейронных сетей // Инновационные кластеры цифровой экономики: драйверы развития. Труды научно-практической конференции с международным

- участием. Под редакцией А.В. Бабкина. – 2018. – С. 481-487.
- [11] Андреев К. В., Быков А. А., Киселева О. М. Математическая модель предиктивного кодирования радиотехнических сигналов, основанная на алгоритме изменяющегося шага кодирования // *Современные наукоемкие технологии*. 2020. – № 11-2. – С. 261-267.
- [12] Андриевская Н. В. Идентификация нелинейной модели с использованием моделей нечеткой логики и искусственных нейронных сетей // *Нейрокомпьютеры: разработка, применение*. – 2017. – № 6. – С. 3-8.
- [13] Зоткина А. А. Рекуррентные нейронные сети как алгоритм последовательности данных // *Современные информационные технологии*. – 2022. – № 35 (35). – С. 24-26.
- [14] Куров А. С., Николаева И. В. Рекуррентные нейронные сети как инструмент прогнозирования временных рядов // *Информационное общество: современное состояние и перспективы развития. Сборник материалов XII международного форума*. – 2019. – С. 272-275.
- [15] Борисенкова А. В., Козлов С. В. Использование метода каскадов Хаара при распознавании образов на изображениях // *Развитие научно-технического творчества детей и молодежи: Сборник материалов III Всероссийской научно-практической конференции с международным участием*. – 2019. – С. 28-33.
- [16] Qi-Qiao He, Cuiyu Wu, Yain-Whar Si LSTM with particle Swam optimization for sales forecasting. *Electronic Commerce Research and Applications*, Volume 51, Jan-Feb 2022. <https://doi.org/10.1016/j.elerap.2022.101118>
- [17] Мошкарлова Л. А., Тельминов О. А. Методы извлечения акустических признаков в задаче распознавания речи рекуррентными нейронными сетями с долгой краткосрочной памятью // *Наноиндустрия*. – 2020. – Т. 13. – № S5-3 (102). – С. 838-841.
- [18] Федотов Д. В., Верхоляк О. В., Карпов А. А. Контекстное непрерывное распознавание эмоций в русской речи с использованием рекуррентных нейронных сетей // *Анализ разговорной русской речи (АРз-2019). Труды восьмого междисциплинарного семинара*. – 2019. – С. 96-99.
- [19] Багаев И. И. Анализ понятий нейронная сеть и сверточная нейронная сеть, обучение сверточной нейросети при помощи модуля Tensorflow // *Математическое и программное обеспечение систем в промышленной и социальной сферах*. – 2020. – Т. 8. № 1. – С. 15-22.
- [20] Захаров В. Н., Мунерман В. И. Параллельный алгоритм умножения многомерных матриц // *Современные информационные технологии и ИТ-образование*. – 2015. – Т. 11. № 2. – С. 384-390.
- [21] Болотова Ю. А., Федотова Л. С., Спицын В. Г. Алгоритм детектирования областей лиц и рук на изображении на основе метода Виолы-Джонса и алгоритма цветовой сегментации // *Фундаментальные исследования*. – 2014. – № 11 - 10. – С. 2130 - 2134.
- [22] Аверкин А. Н., Соболев С. В., Воронцов А. О. Сравнение различных техник анализа эмоций для решения задачи визуализации индекса настроения // *Мягкие измерения и вычисления*. – 2019. – №11. (24). – С. 30-34.
- [23] Козлов С. В., Светлаков А. В. О LL(1)-грамматиках, алгоритмах на них и методах их анализа в программировании // *International Journal of Open Information Technologies*. – 2022. Т. 10. № 3. – С. 30-38.
- [24] Tasarruf Bashir, Chen Haoyong, Muhammad Faizan Tahir, Zhu Liqiang Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. *Energy Reports*, Volume 8, November 2022, Pp. 1678-1686, Energy Reports. <https://doi.org/10.1016/j.egy.2021.12.067>
- [25] Jiaqi Qin, Yi Zhang, Shixiong Fan, Xiaonan Hu, Yongqiang Huang, Zexin Lu, Yan Liu Multi-task short-term reactive and active load forecasting method based on attention-LSTM model. *International Journal of Electrical Power & Energy Systems*, Vol. 135, February 2022. <https://doi.org/10.1016/j.ijepes.2021.107517>
- [26] Shiva Nosouhian, Fereshteh Nosouhian, Abbas Kazemi Khoshouei A review of recurrent neural network architecture for sequence learning: comparison between LSTM and GRU. *Preprints 2021*, 2021070252 <https://doi.org/10.20944/preprints202107.0252.v1>
- [27] Xu G., Peng Sh., Li Ch., Chen X. Synergistic evolution of China's green economy and digital economy based on LSTM-GM and grey absolute correlation // *Sustainability*. 2023. Vol. 15. № 19. P. 14156.
- [28] Ezat Ahmadzadeh, Hyunil Kim, Ongee Jeong, Namki Kim, Inkyu Moon A deep bidirectional lstm-gru network model for automated ciphertext classification. *IEEE Access*, Vol. 10. P. 3228-3237.
- [29] Савченко В. В. Метод авторегрессионного моделирования речевого сигнала с использованием огибающей периодограммы Шустера в качестве опорного спектрального образца // *Радиотехника и электроника*. – 2023. – Т. 68. № 2. – С. 138-145.
- [30] Ya Gao, Rong Wang, Enmin Zhou Stock prediction based on optimized LSTM and GRU models. *Hindawi, Scientific Programming*, Volume 2021. <https://doi.org/10.1155/2021/4055281>

LSTM and GRU model analysis for time series forecasting

S.V. Kozlov, S.A. Sedenkov

Abstract – The article presents an analysis of the application of recurrent neural networks as a tool for predicting time series. General principles of operation of recurrent neural networks are briefly described. The advantages of their use in comparison with standard neural networks and convolutional networks are described. Application areas of different types of recurrent neural network architecture are considered. The algorithm of functioning of recurrent neural networks was analyzed. The main class is defined in the description of the algorithm, the software implementation of its functions is given. Particular attention is paid to the matrix form of the parameters when executing the algorithm. The main part of the work is devoted to a comparative analysis of the model of a long chain of short-term memory elements and the model of controlled recurrent blocks. The article briefly describes the history of their development, the principle of operation of each model. The possibilities of these models in solving problems of constructing time series are considered. Disclosed is the essence of formulas, with the help of which neural networks LSTM and GRU perform calculations. The program code developed by the authors for each of the models is characterized. IBM stock price forecasting was chosen to analyze the operation of algorithms. The data obtained during the experiment are shown in the graphs. At the end of the work, their comparative analysis is given. The relevance of the article is due to the effectiveness of the implementation of methods for recurrent analysis of time series data using neural networks.

Keywords – recurrent neural network, long chain of short-term memory elements, controlled recurrent blocks, modeling, quotation cost forecasting, program, neural network technologies, information technologies.

REFERENCES

- [1] Antonov V. V., Pal'chevskij E. V., Enikeev R. R. Prognozirovanie na osnove iskusstvennoj nejronnoj seti vtorogo pokolenija dlja podderzhki prinjatija reshenij v osobo znachimyh situacijah // Programmnye produkty i sistemy. – 2022. – # 3. – S. 384-395.
- [2] Kozlov S. V. Ispol'zovanie funkcional'nyh vozmozhnostej informacionnyh sistem v proizvodstvennoj sfere // JeNERGETIKA, INFORMATIKA, INNOVACII – 2017 (jelektrojenergetika, jelektrotehnika i teplojenergetika, matematicheskoe modelirovanie i informacionnye tehnologii v proizvodstve). Sbornik trudov VII-oj Mezhdunarodnoj nauchno-tehnicheskoy konferencii. – 2017. – V 3 t. T 1. – S. 298-301.
- [3] Fahrutdinova A. Z. Metodologicheskie problemy sovremennoj prognostiki // Omskie nauchnye chtenija. Materialy Vserossijskoj nauchno-prakticheskoy konferencii. – 2017. – S. 851-853.
- [4] Kozlov S. V., Suin I. A. O nekotoryh aspektah primeneniya invariantnyh metodov funkcional'nogo analiza dannyh v razlichnyh predmetnyh oblastjah // Sistemy komp'juternoj matematiki i ih prilozhenija. – 2019. – # 20-1. – S. 199-205.
- [5] Dubenko Ju. V., Dyshkant E. E. Nejrosetevoj algoritm vybora metodov dlja prognozirovanija vremennyh rjadov // Vestnik Astrahanskogo gosudarstvennogo tehnicheskogo universiteta. Serija: Upravlenie, vychislitel'naja tehnika i informatika. – 2019. # 1. S. 51-60.
- [6] Kozlov S. V. Ispol'zovanie sootvetstvija Galua kak invarianta otbora kontenta pri proektirovanii informacionnyh sistem // Sovremennye informacionnye tehnologii i IT-obrazovanie. – 2015. – T. 2. # 11. – S. 220-225.
- [7] Alymova E. V. Sovmestnoe primenenie modeli linejnoj regressii i nejronnoj seti v zadache predskazanija trenda kotirovok kriptovaljuty Bitcoin // Inzhenernyj vestnik Dona. – 2020. – # 10 (70). – S. 90-96.
- [8] Senotova S. A. Sravnitel'nyj analiz metodov approksimacii s pomoshh'ju regressionnyh zavisimostej i nejronnyh setej dlja linejnyh modelej // Sbornik nauchnyh trudov Angarskogo gosudarstvennogo tehnicheskogo universiteta. – 2021. – T. 1. # 18. – S. 31-35.
- [9] Solov'eva E. B. Rekurrentnye nejronnye seti v kachestve modelej nelinejnyh dinamicheskikh sistem // Cifrovaja obrabotka signalov. – 2018. – # 1. – S. 18-27.
- [10] Dubolazov V. A., Somov A. G. Sozdanie nelinejnyh approksimacij sovremennyh jekonomicheskikh modelej metodom nejronnyh setej // Innovacionnye klasteri cifrovoj jekonomiki: drajvery razvitija. Trudy nauchno-prakticheskoy konferencii s mezhdunarodnym uchastiem. Pod redakciej A.V. Babkina. – 2018. – S. 481-487.
- [11] Andreev K. V., Bykov A. A., Kiseleva O. M. Matematicheskaja model' prediktivnogo kodirovanija radiotehnicheskikh signalov, osnovannaja na algoritme izmenjajushhegosja shaga kodirovanija // Sovremennye naukoemkie tehnologii. 2020. – # 11-2. – S. 261-267.
- [12] Andrievskaja N. V. Identifikacija nelinejnoj modeli s ispol'zovaniem modelej nechetkoj logiki i iskusstvennyh nejronnyh setej // Nejrokomp'jutery: razrabotka, primenenie. – 2017. – # 6. – S. 3-8.
- [13] Zotkina A. A. Rekurrentnye nejronnye seti kak algoritm posledovatel'nosti dannyh // Sovremennye

informacionnye tehnologii. – 2022. – # 35 (35). – S. 24-26.

[14] Kurov A. S., Nikolaeva I. V. Rekurrentnye nejronnye seti kak instrument prognozirovanija vremennyh rjadov // Informacionnoe obshchestvo: sovremennoe sostojanie i perspektivy razvitija. Sbornik materialov XII mezhdunarodnogo foruma. – 2019. – S. 272-275.

[15] Borisenkova A. V., Kozlov S. V. Ispol'zovanie metoda kaskadov Haara pri raspoznavanii obrazov na izobrazhenijah // Razvitie nauchno-tehnicheskogo tvorcestva detej i molodezhi: Sbornik materialov III Vserossijskoj nauchno-prakticheskoy konferencii s mezhdunarodnym uchastiem. – 2019. – S. 28-33.

[16] Qi-Qiao He, Cuiyu Wu, Yain-Whar Si LSTM with particle Swam optimization for sales forecasting. Electronic Commerce Research and Applications, Volume 51, Jan-Feb 2022. <https://doi.org/10.1016/j.elerap.2022.101118>

[17] Moshkarova L. A., Tel'minov O. A. Metody izvlechenija akusticheskikh priznakov v zadache raspoznavanija rechi rekurrentnymi nejronnymi setjami s dolgoj kratkosrochnoj pamjat'ju // Nanoindustrija. – 2020. – T. 13. – # S5-3 (102). – S. 838-841.

[18] Fedotov D. V., Verholjak O. V., Karpov A. A. Kontekstnoe nepreryvnoe raspoznavanie jemocij v ruskoj rechi s ispol'zovaniem rekurrentnyh nejronnyh setej // Analiz razgovornoj ruskoj rechi (ARz-2019). Trudy vos'mogo mezhdisciplinarnogo seminaru. – 2019. – S. 96-99.

[19] Bagaev I. I. Analiz ponjatij nejronnaja set' i svertochnaja nejronnaja set', obuchenie svertochnoj nejroseti pri pomoshhi modulja Tensorflow // Matematicheskoe i programmnoe obespechenie sistem v promyshlennoj i social'noj sferah. – 2020. – T. 8. # 1. – S. 15-22.

[20] Zaharov V. N., Munerman V. I. Parallelnyj algoritm umnozhenija mnogomernyh matric // Sovremennye informacionnye tehnologii i IT-obrazovanie. – 2015. – T. 11. # 2. – S. 384-390.

[21] Bolotova Ju. A., Fedotova L. C., Spicyn V. G. Algoritm detektirovanija oblastej lic i ruk na izobrazhenii na osnove metoda Violy-Dzhonsa i algoritma cvetovoj

segmentacii // Fundamental'nye issledovanija. – 2014. – # 11 - 10. – S. 2130 - 2134.

[22] Averkin A. N., Sobolev S. V., Voroncov A. O. Sravnenie razlichnyh tehnik analiza jemocij dlja reshenija zadachi vizualizacii indeksa nastroenija // Mjagkie izmerenija i vychislenija. – 2019. – #11. (24). – S. 30-34.

[23] Kozlov S. V., Svetlakov A. V. O LL(1)-grammatikah, algoritmah na nih i metodah ih analiza v programmirovanii // International Journal of Open Information Technologies. – 2022. T. 10. # 3. – S. 30-38.

[24] Tasarruf Bashir, Chen Haoyong, Muhammad Faizan Tahir, Zhu Liqiang Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. Energy Reports, Volume 8, November 2022, Pp. 1678-1686, Energy Reports. <https://doi.org/10.1016/j.egy.2021.12.067>

[25] Jiaqi Qin, Yi Zhang, Shixiong Fan, Xiaonan Hu, Yongqiang Huang, Zexin Lu, Yan Liu Multi-task short-term reactive and active load forecasting method based on attention-LSTM model. International Journal of Electrical Power & Energy Systems, Vol. 135, February 2022. <https://doi.org/10.1016/j.ijepes.2021.107517>

[26] Shiva Nosouhian, Fereshteh Nosouhian, Abbas Kazemi Khoshouei A review of recurrent neural network architecture for sequence learning: comparison between LSTM and GRU. Preprints 2021, 2021070252 <https://doi.org/10.20944/preprints202107.0252.v1>

[27] Xu G., Peng Sh., Li Ch., Chen X. Synergistic evolution of China's green economy and digital economy based on LSTM-GM and grey absolute correlation // Sustainability. 2023. Vol. 15. # 19. P. 14156.

[28] Ezat Ahmadzadeh, Hyunil Kim, Ongee Jeong, Namki Kim, Inkyu Moon A deep bidirectional lstm-gru network model for automated ciphertext classification. IEEE Access, Vol. 10. P. 3228-3237.

[29] Savchenko V. V. Metod avtoregressionnogo modelirovanija rechevogo signala s ispol'zovaniem ogibajushhej periodogrammy Shustera v kachestve opornogo spektral'nogo obrazca // Radiotehnika i jelektronika. – 2023. – T. 68. # 2. – S. 138-145.

[30] Ya Gao, Rong Wang, Enmin Zhou Stock prediction based on optimized LSTM and GRU models. Hindawi, Scientific Programming, Volume 2021. <https://doi.org/10.1155/2021/4055281>