

ИССЛЕДОВАНИЕ ВАРИАНТОВ АРХИТЕКТУР НЕЙРОННЫХ СЕТЕЙ С ПАМЯТЬЮ НА ОСНОВЕ НЕЧЕТКОЙ ЛОГИКИ В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДОВ

КУМСКОВ МИХАИЛ ИВАНОВИЧ (KUMSKOV MIKHAIL IVANOVICH)^a
д.ф.-м.н., профессор кафедры Вычислительной математики Механико-математического
факультета МГУ им. М.В.Ломоносова, заведующий лабораторией вычислительных методов
e-mail: mikhail.kumskov@math.msu.ru

МАХОВА АНАСТАСИЯ ГЕННАДЬЕВНА (MAKHOVA ANASTASIIA GENNADEVNA)^a
e-mail: anastasiiia.makhova@math.msu.ru

^aФедеральное государственное бюджетное образовательное учреждение высшего
образования «Московский государственный университет имени М. В. Ломоносова», Россия,
Москва, тер. Ленинские Горы, д. 1, 119991

Аннотация В данной работе рассматриваются архитектуры нейронных сетей с памятью, использующие принципы нечеткой логики для прогнозирования финансовых временных рядов. Нечеткая логика применяется как на этапе предобработки данных, так и внутри модели, обеспечивая гибкое представление входных характеристик для более точного прогноза.

Прогнозирование финансовых временных рядов представляет собой сложную задачу из-за высокой волатильности и наличия нелинейных зависимостей с историческими данными. Для ее решения используется Long Short-Term Memory — модель, зарекомендовавшая себя в работе с временными рядами.

На этапе предобработки применяется фаззификация входных данных, позволяющая представить их в виде нечетких множеств. Это способствует снижению чувствительности модели к шуму и выделению значимых паттернов в данных.

Внутри модели фаззификация осуществляется через функции активации, которые рассматриваются как функции принадлежности к нечетким множествам. Это позволяет нейросети проводить операции нечеткой логики, регулируя наклон и смещение точки нуля функций активации. Тем самым достигается адаптивная настройка модели под специфику конкретных финансовых данных, что позволяет лучше улавливать сложные зависимости, снижая влияние шума в исходном ряде.

Результаты вычислительных экспериментов показывают, что предложенный подход демонстрирует лучшую предсказательную способность, чем стандартная Long Short-Term Memory -модель.

Ключевые слова: Финансовые временные ряды, функция активации, нечеткая логика, нейронные сети с памятью

ВВЕДЕНИЕ

Прогнозирование финансовых временных рядов — одна из ключевых задач в экономике и финансах. Оно применяется в алгоритмической торговле, управлении инвестициями, оценке рисков и макроэкономическом анализе. Однако такие особенности финансовых временных рядов как нестационарность, наличие нелинейных зависимостей и шума [15] (А.Н., 2016) делают эту задачу сложной для традиционных статистических моделей, таких как ARIMA и GARCH [2].

В последние годы значительное внимание привлекают методы глубокого обучения, особенно рекуррентные нейросети (RNN) и их усовершенствованные версии, такие как нейросети с долгой краткосрочной памятью (Long Short-Term Memory, LSTM)[14]. Эти модели способны улавливать сложные временные зависимости и эффективно работать с временными рядами. Однако стандартные LSTM-модели чувствительны к шуму в данных и используют фиксированные функции активации, что ограничивает их адаптивность к специфике временного ряда.

В данной работе предлагается использование нечеткой логики (fuzzy logic) для улучшения прогнозирования финансовых временных рядов. Нечеткая логика позволяет гибко обрабатывать данные, учитывая их неопределенность и варьированность. Ее применение возможно на нескольких этапах работы нейросети:

Фаззификация входных данных — преобразование числовых значений в нечеткие множества, что позволяет сглаживать шум. В работе [1] рассматриваются варианты представления нечетких временных рядов: распределение нечеткости, моделируемое лингвистической переменной, распределение нечеткости во времени или по переменной значения. Колеблющиеся нечеткие множества (Hesitant Fuzzy Set (HFS)) применяются для прогнозирования баллов для поступления в университет Алабамы и цен акций индийских компаний в работе [7].

Использование нечетких функций активации -- изменение наклона и сдвига функций активации, что позволяет подстраиваться под особенности конкретного временного ряда. В предыдущей работе авторов данной статьи [8] было показано, что функции активации *cloglog* и *loglog*, которые можно рассмотреть как преобразования сигмоидальной функции

активации, улучшают прогноз. В литературе встречаются свежие работы о параметрических [4] и нечетких [3] функциях активации.

Основной целью работы является исследование архитектур нейронных сетей с памятью, использующих принципы нечеткой логики для повышения точности прогнозирования финансовых временных рядов. Для достижения этой цели в работе решаются следующие задачи: изучение теоретических основ нечеткой логики и ее применения в машинном обучении, разработка и тестирование метода фаззификации входных данных, внедрение модифицированных функций активации на основе нечеткой логики в LSTM-сеть, сравнительный анализ точности предложенного метода с классическими моделями (стандартной LSTM).

Статья организована следующим образом. В разделе 1 рассматриваются основные теоретические аспекты, необходимые в работе: принципы нечеткой логики, роль функции активации, особенности финансовых временных рядов. В разделе 2 приводится постановка задачи. Принцип работы LSTM описан в разделе 3. В разделе 4 отмечены варианты функций потерь, алгоритмов оптимизации, метрик оценки качества, которые подходят для решения задачи прогнозирования финансовых временных рядов. В разделе 5 представлена разработанная модель, приведены результаты вычислительных экспериментов и их анализ. В Заключение подводятся итоги исследования и рассматриваются возможные направления дальнейшей работы.

1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И НОТАЦИЯ

1.1 Финансовые временные ряды

Определение 1.1 Временной ряд X представляет собой последовательность хронологически упорядоченных вещественных векторов:

$$X = (x_{t_1}, x_{t_2}, \dots, x_{t_n}), \quad x_{t_i} \in \mathbb{R}^m.$$

Число n обозначается как $|X|$ и называется **длиной ряда**.

Постоянный шаг по времени назовем **тиком**: $\Delta t = t_i - t_{i-1}$

Определение 1.2 Финансовый временной ряд имеет ряд особенностей, в частности, нелинейный характер формирования значений финансовых индексов и последствие, выражающееся в том, что многие индексы, цены и тд. "помнят" прошлое. [15]

1.2. Задачи прогнозирования финансовых временных рядов

Задачи о прогнозировании финансовых временных рядов можно разделить на 2 группы в зависимости от ожидаемых выходных данных [14]:

- **прогноз цены** – необходимо предсказать стоимость на некоторое время вперед.

- **прогноз тренда** – направления движения графика стоимости:
 - 2-class problem – предсказать восходящий и нисходящий тренд
 - 3-class problem - предсказать восходящий, нисходящий и боковой тренды

Входные данные для прогнозирования можно брать из:

- Цен за предыдущие периоды.
- Фундаментального анализа. Благодаря методам фундаментального анализа можно проанализировать справедливую стоимость на данный момент и предсказать повышение/понижение спроса
- Технического анализа. Анализ графиков на характерные предпосылки к изменению направления движения стоимости.
- Текстовой информации. Обработка новостей и определение эмоционального настроения и мнений на основе постов и комментариев в социальных сетях помогают спрогнозировать поведение инвесторов.

1.3. Роль функции активации

Основа искусственных нейронных сетей — это *нейрон* и *функция активации*.

Определение 1.3 Нейрон – элемент, который принимает, обрабатывает и передает сигнал другим элементам нейросети.

Определение 1.4 Функция активации – это математическая функция, которая по нескольким входным сигналам определяет один выходной сигнал.

Функция активации определяет, как нейрон будет реагировать на входные данные: для входов вычисляется взвешенная сумма со смещением, выполняется преобразование с помощью функции активации, и в результате получается выходной сигнал.

$$y_k = \phi_k \left(\sum_i w_{ij} * x_i + b_k \right)$$

Где x_i - входной сигнал, y_j —выходной сигнал, ϕ - функция активации, w_{ij} - весовые коэффициенты и b_j – смещение.

Каждый нейрон по сути выполняет нелинейное преобразование входного сигнала в выходной по формуле выше. Саму нейросеть можно рассматривать как композицию взвешенных нейронов [5].

В работе упоминаются s-образные функции активации: *tanh, sigmoid, cloglog, loglog*. Вид функций активации и их области значений приведены в таблице 1. Графики функций активаций отображены на рисунке 1.

Функция активации	$\phi(y) =$	$\phi'(y) =$	Область значений
<i>cloglog</i>	$1 - \exp(-\exp(y))$	$\exp(y * (1 - \phi(y)))$	$(0; 1)$
<i>loglog</i>	$\exp(-\exp(-y))$	$\exp(-y * (1 - \phi(y)))$	$(0; 1)$
<i>tanh</i>	$\frac{\exp(2y) - 1}{\exp(2y) + 1}$	$1 - \phi^2(y)$	$(-1; 1)$
<i>sigmoid</i>	$\frac{1}{1 + \exp(-y)}$	$\phi(y) * (1 - \phi(y))$	$(0; 1)$

Таблица 1. Функции активации, их первые производные и области значений.

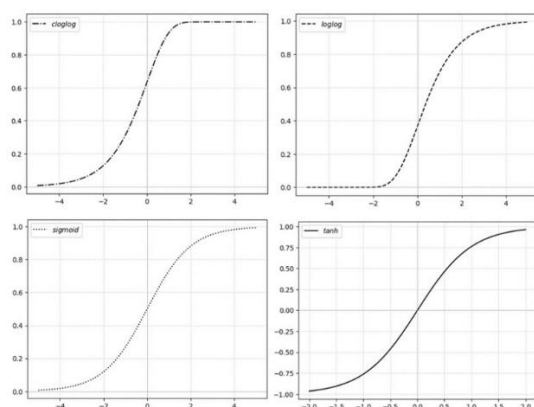


Рисунок 1. Графики функций активации: *cloglog*, *loglog*, *tanh*, *sigmoid*

1.4. Нечеткая логика

Определение 5. Нечеткое множество (fuzzy set) - это пара (U, m) , где U - универсальное множество, $m: U \rightarrow [0, 1]$ - **функция принадлежности** (membership function). Для каждого $x \in U$ величина $m(x)$ называется **степенью принадлежности** x к нечеткому множеству (U, m) . Для конечного $U = \{x_1, \dots, x_n\}$ нечеткое множество (U, m) обозначается как

$$\{m(x_1)/x_1, \dots, m(x_n)/x_n\}$$

Пусть $x \in U$, тогда:

$$m(x) = \begin{cases} 0 & x \text{ не лежит в нечетком множестве} \\ 1 & x \text{ полностью принадлежит нечеткому множеству (full member)} \\ \text{иначе} & x \text{ частично принадлежит нечеткому множеству (fuzzy member)} \end{cases}$$

1.4.1 Операции над нечеткими множествами

Пересечение двух нечетких множеств (нечеткое «И»): $m_{A \cap B}(x) = \min(m_A(x), m_B(x))$

Объединение двух нечетких множеств (нечеткое «ИЛИ»): $m_{A \cup B}(x) = \max(m_A(x), m_B(x))$

1.4.2 Переменные

Для описания нечетких множеств вводятся понятия нечеткой и лингвистической переменных.

Определение 6. **Нечеткая переменная** описывается набором (N, U, A) , где N — это название переменной, U — универсальное множество (область рассуждений), A — нечеткое множество на U .

Значениями **лингвистической переменной** могут быть нечеткие переменные, т.е. лингвистическая переменная находится на более высоком уровне, чем нечеткая переменная. Каждая лингвистическая переменная состоит из: названия; множества своих значений, которое также называется базовым терм-множеством T . Элементы базового терм-множества представляют собой названия нечетких переменных; универсального множества U ; синтаксического правила G , по которому генерируются новые термы с применением слов естественного или формального языка; семантического правила P , которое каждому значению лингвистической переменной ставит в соответствие нечеткое подмножество множества U .

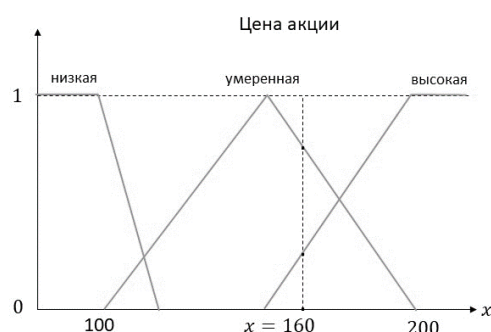


Рис. 2 Пример лингвистической переменной “Цена акции”. Нечеткие переменные: “низкая”, “умеренная” и “высокая”. Универсальное множество $U = [100, 200]$. Степень принадлежности цены 160 к множеству “низкая” - нулевая т.е. $m_{\text{низкая}}(160) = 0$, для множества “умеренная” $m_{\text{умеренная}}(160) = 0.75$ и для множества “высокая” $m_{\text{высокая}}(160) = 0.40$.

1.4.3 Функции принадлежности

Существует свыше десятка типовых форм кривых для задания функций принадлежности. Наибольшее распространение получили: треугольная, трапецеидальная и гауссова функции принадлежности.

Треугольная функция принадлежности определяется тройкой чисел (a, b, c) . Аналогично для задания трапецеидальной функции принадлежности необходима четверка чисел (a, b, c, d) .

Значения функций в точке x вычисляются согласно выражениям: $m_{\text{треугольная}}(x) =$

$$\begin{cases} 1 - \frac{b-x}{b-a} & a \leq x \leq b \\ 1 - \frac{x-b}{c-b} & b \leq x \leq c \\ 0 & x \notin (a; c) \end{cases}, \quad m_{\text{трапецеидальная}}(x) = \begin{cases} 1 - \frac{b-x}{b-a} & a \leq x \leq b \\ 1 & b \leq x \leq c \\ 1 - \frac{x-b}{c-b} & c \leq x \leq d \\ 0 & x \notin (a; d) \end{cases}$$

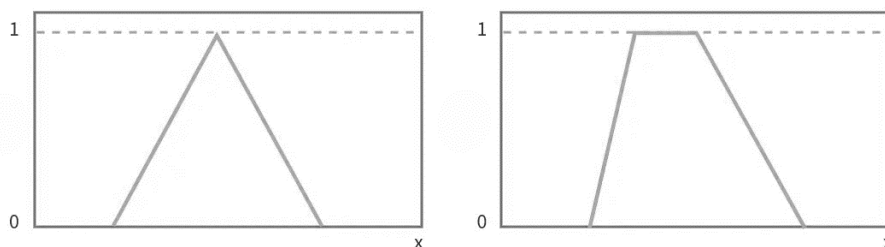


Рис. 3. Слева: треугольная функция активации.

Справа: трапецеидальная функция активации

Функция принадлежности гауссова типа описывается формулой: $m(x) = \exp(-(\frac{x-c}{\sigma})^2)$

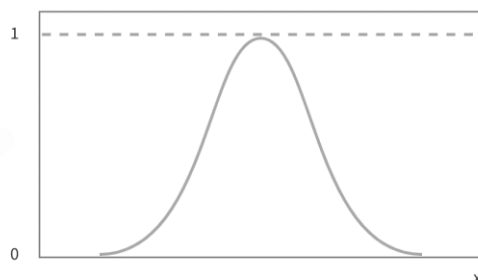


Рис. 4. Функция принадлежности гауссова типа.

Кроме того, функцией принадлежности могут быть и функции активации, чьи значения лежат в $[0,1]$, например, s-образные функции *sigmoid*, *cloglog*, *loglog* и их вариации со смещенным центром нуля и измененным наклоном.

1.4.4 Нечеткий логический вывод

Основой для проведения операции нечеткого логического вывода является *база правил*, содержащая нечеткие высказывания и функции принадлежности для соответствующих лингвистических термов.

Механизм логического вывода включает четыре этапа: введение нечеткости (**фаззификация**), нечеткий вывод, композиция и приведение к четкости, или **дефаззификация**. Алгоритмы нечеткого вывода различаются, главным образом, видом

используемых правил, логических операций и разновидностью метода дефаззификации. Известны модели нечеткого вывода Мамдани, Сугено, Ларсена, Цукамото. [12]



Рис. 5. Механизм логического вывода. На вход подаются четкие данные X , они фаззифицируются(приводятся к нечеткому виду \tilde{X}). Нечеткий логический вывод производится с помощью некой базы правил. Результат \tilde{Y} дефаззифицируется (приводится к четкому виду Y)

2 ПОСТАНОВКА ЗАДАЧИ

Нам дан одномерный финансовый временной ряд цен

$$X = (x_{t_1}, x_{t_2}, \dots, x_{t_n}), \quad x_{t_i} \in \mathbb{R}$$

с тиком $\Delta t = \{\text{минута, час, день}\}$

Нужно спрогнозировать значения ряда на k тиков вперед: $Y = (y_{t_{n+1}}, y_{t_{n+2}}, \dots, y_{t_{n+k}})$

Общий вид модели будет напоминать схему логического вывода, где за нечеткий логический вывод будет отвечать нейронная сеть с модифицированными функциями активации, которая будет производить операции над нечеткими множествами с функциями принадлежности отождествленными с функциями активации.

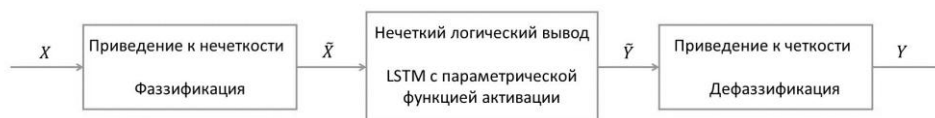


Рис. 6. Общая схема предложенной модели. На вход подаются четкие данные X , они фаззифицируются(приводятся к нечеткому виду \tilde{X}). Нечеткий логический вывод производится LSTM с параметрическими функциями активации. Результат \tilde{Y} дефаззифицируется (приводится к четкому виду Y).

2.1 Фаззификация

Ввести нечеткость в финансовый временной ряд можно несколькими способами:

Гауссианы. Для каждого момента времени t_i сопоставим известному значению ряда x_i нечеткое множество, задающееся функцией принадлежности гауссовского типа

$$m(x) = \exp\left(-\left(\frac{x - c}{\sigma}\right)^2\right)$$

с параметрами $c = x_i$ и $\sigma = \hat{\sigma}(t_i, l)$ – волатильность ряда за период длины l : $[t_{i-l}, t_i]$. То есть вершина гауссовского купола будет как раз в x_i и его “ширина” будет соответствовать “ширине” обзореваемого ряда.

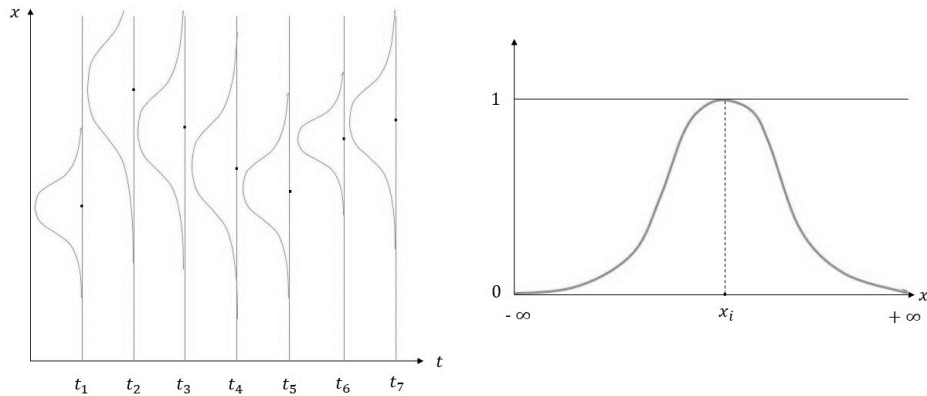


Рис. 7. Слева: точки – значение временного ряда x_i , гауссовские “купола” – условное обозначение нечеткого временного ряда: каждому моменту времени t_i сопоставлено нечеткое множество. Справа “фотография” в фиксированный момент t соответствующего нечеткого множества.

Probabilistic Fuzzy Sets. Этот метод описан в работе [10].

1. Определим область рассуждений $U = [X_{min} - \sigma, X_{max} + \sigma]$, X_{min} и X_{max} – минимальные и максимальные значения ряда за обзореваемый период, σ – дисперсия.
2. Разобьем U на m равных интервалов e_i и зададим на них треугольными функциями принадлежности нечеткие множества F_i

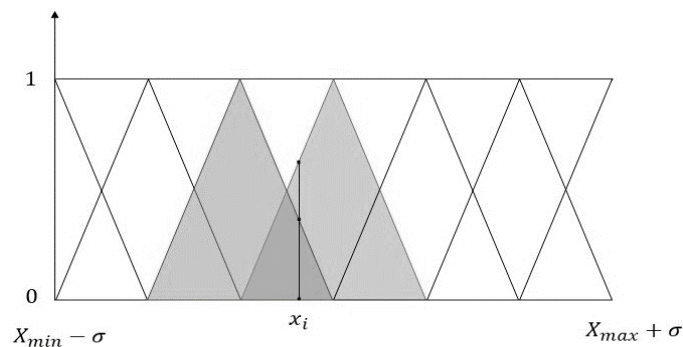


Рис. 8. Разбили U на отрезки e_j и на них, через треугольные функции принадлежности, построим нечеткие множества F_j .

Лингвистическая переменная. Можем перейти от задачи прогнозирования цены к задаче классификации: прогнозированию уровня цены (см. пример на рис. 2), тренда или даже эмоционального настроения участников рынка. В этом случае через треугольные и трапециевидные функции принадлежности зададим по соответствующим лингвистическим переменным нечеткие множества.

2.2 Нечеткость и функция активации

Пусть $\phi_{\theta}(y)$ - это функция активации с параметрами θ , которые определяют форму функции активации (например, точка нуля, наклон). Наша задача – найти оптимальные параметры функции активации θ^* , минимизирующие ошибку прогноза финансового временного ряда:

$$\theta^* = \operatorname{argmin}_{\theta} \mathcal{L}(X, y, \theta)$$

где $\mathcal{L}(X, y, \theta)$ – функция потерь, показывающая как хорошо нейросеть с параметрами θ (т.е. параметрами функции активации в том числе) прогнозирует y по входным данным X [].

Как было отмечено выше, s-образные функции активации такие как *sigmoid* и ее вариации со смещенным центром нуля и измененным наклоном можно рассматривать как функции принадлежности к нечеткому множеству. А так как нейронная сеть — это композиция нейронов производящих вычисления с помощью функций активации, можно сказать, что нейросеть производит операции над нечеткими множествами и является моделью нечеткого вывода.

2.3 Дефаззификация

Приведем нечеткие результаты работы нейронной сети к четким прогнозам.

Гауссианы. Возьмем за прогноз y_i вершину спрогнозированного купола c

Probabilistic Fuzzy Sets. Зная треугольные функции принадлежности на каждом интервале и, спрогнозировав распределение степеней принадлежности к множествам, можем однозначно определить четкий прогноз.

Лингвистическая переменная. Этот случай аналогичен.

3. LSTM

Любая рекуррентная нейронная сеть имеет форму цепочки повторяющихся модулей нейронной сети. (рис. 3) LSTM (Long short-term memory) – тип рекуррентной нейросети [13], модули, которой содержат четыре слоя, вместо одного (рис.4), что помогает эффективно

работать с длинными временными периодами: распознавать долгосрочные закономерности, определять события, находящиеся во времени далеко друг от друга, извлекать информацию, передаваемую расстоянием между событиями. [6]

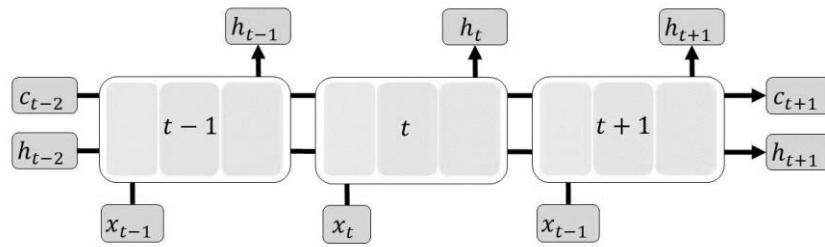


Рис. 9. Структура LSTM, развернутая во времени, напоминает цепочку

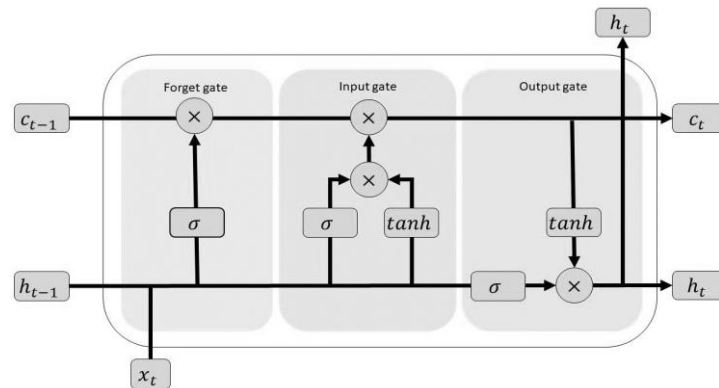


Рис. 10. Архитектура LSTM. Обозначения: c_* —состояние ячейки, x_* —вход, h_* —выход, \times — поточечное умножение, $+$ — сложение векторов, σ — сигмоидальный слой, \tanh — \tanh -слой.

Для описания работы архитектуры нейронной сети LSTM для прогнозирования введем несколько основных компонент и операций. Для удобства обозначим временной ряд цен как $\{x^{(1)}, x^{(2)}, \dots, x^{(T)}\}$, где $x^{(t)}$ - это цена биткоина в момент времени t . Для каждого момента времени t , входной вектор $x^{(t)}$ содержит информацию о цене.

Нейронная сеть LSTM состоит из нескольких ячеек LSTM, каждая из которых имеет вход $x^{(t)}$, предыдущее скрытое состояние $c^{(t-1)}$ и предыдущий выход $h^{(t-1)}$.

Ключевой компонент LSTM – это состояние ячейки (cell state) C_* . LSTM может удалять информацию из состояния ячейки; этот процесс регулируется структурами, называемыми фильтрами (gates). В LSTM три таких фильтра, позволяющих определять какая информация должна быть сохранена или забыта. Фильтры состоят из слоя сигмоидальной (σ) нейронной сети и операции поточечного умножения. Сигмоидальный слой возвращает числа от нуля до единицы, которые обозначают, какую долю каждого блока информации следует пропустить дальше по сети. Ноль в данном случае означает “не пропускать ничего”, единица – “пропустить все”. [11]

Простейший LSTM-модуль можно представить в виде системы уравнений:

$$\begin{cases} f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_C * [h_{t-1}, x_t] + b_C) \\ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \\ h_t = o_t * \tanh(C_t) \end{cases}$$

3 МОДЕЛЬ

Ключевые компоненты любой нейросети: функции потерь, алгоритмы оптимизации, метрики оценки качества и функции активации. Они определяют как нейронная сеть будет обучаться, как мы будем интерпретировать результаты. Выбор каждого из методов зависит от задачи, типа данных и отображается на правильности и скорости работы алгоритма.

В случае прогнозирования финансовых временных рядов стоит выбрать:

1. Функция потерь: MSE
2. Алгоритм оптимизации: Adam – это метод оптимизации, который обновляет веса модели на основе градиентов и квадратов градиентов. Это один из самых популярных оптимизаторов благодаря своей скорости сходимости и стабильности.
3. Метрики оценки качества прогноза: MAE, RMSE, MAPE – метрики для задач регрессии. С их помощью можем оценить на тестовом множестве качество предсказаний обученной нейросети, отследить переобучение.

6. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

6.1 Задачи для вычислительного эксперимента

В рамках вычислительного эксперимента необходимо решить следующие задачи:

1. Исследовать влияние фаззификации входных данных на точность прогнозирования временного ряда.
2. Исследовать влияние параметрических функций активации внутри LSTM на качество прогнозирования временного ряда.
3. Провести сравнение результатов предложенной модели с традиционной моделью LSTM.

6.2 Описание вычислительного эксперимента

6.2.1 Сбор данных

Данные для экспериментов были получены с финансовой платформы Yahoo Finance. Для анализа и прогнозирования использовались данные котировок криптовалюты BTC-USD (биткоин к доллару США) с разными интервалами: 1 минута, 1 час и 1 день. Период, охваченный экспериментами:

- 1-минутный интервал: последняя неделя.
- 1-часовой интервал: последние 730 дней
- 1-дневной интервал: вся история наблюдений

Полученные данные были предварительно обработаны с помощью нормализации (и фаззификации), а затем разделены на тренировочные (75%) и тестовые (25%) выборки.

Диапазон значений исходного временного ряда составляет от 81,000 до 88,000 долларов. Это важно учитывать при интерпретации метрик качества прогноза.

6.2.2 Проведение эксперимента

Для обучения возьмем число эпох = 100, размер батча = 32. Все модели запускались, в зависимости от тика, на следующих параметрах:

Тик	units	lookback
1m	16	30
1h	16	30
1d	16	1

Таблица 2. Параметры units и lookback в зависимости от тика.

Сравним результаты работы следующих архитектур:

LSTM

– базовая модель LSTM без применения фаззификации и параметрической активации. Размерность входных данных: (*lookback*, 1). Состоит из слоя *LSTM(units)* и полносвязного слоя *Dense(1)*.

Fuzzy Gauss

– LSTM с добавленной фаззификацией данных методом [гауссиан](#). Локальную волатильность (“ширину” купола) будем считать по *window* = 10 последним измерениям. Размерность

входных данных: $(lookback, 2)$. Модель состоит из слоя $LSTM(units)$ и полносвязного слоя $Dense(2)$.

PFS

– LSTM с добавленной фазификацией данных с помощью [Probabilistic Fuzzy Sets](#). Число интервалов $m = 14$, как в оригинальной статье. Размерность входных данных: $(lookback, m)$. Состоит из слоя $LSTM(units)$ и полносвязного слоя $Dense(m)$.

Fuzzy Activation

– LSTM с параметрическими функциями активации. Есть несколько вариантов:

1. Каждая из трех сигмOID в LSTM имеют одинаковую параметризацию. (+2 параметра)
2. Все три сигмOIDы в LSTM имеют разные параметры. (+6 параметров)
3. Параметризуем все 5 функций активации в LSTM. Заменяем \tanh на параметрическую сигмOIDу. (+10 параметров)

Каждое увеличение количества параметров модели увеличивает время обучения. Размерность входных данных: $(lookback, 1)$. Состоит из слоя $LSTM(units)$ и полносвязного слоя $Dense(1)$.

Fuzzy LSTM

– итоговая модель с введением нечеткости на этапе предобработки данных и внутри модели. Для эксперимента выбрана комбинация лучших по качеству прогноза версий моделей выше: Fuzzy Gauss + Fuzzy Activation 1 с одной параметрической сигмOIDой.

6.3 Результаты вычислительного эксперимента

6.3.1 Минутные тики

Результаты эксперимента на минутных тиках представлены в таблице [1](#). Качество прогнозирования оценивалось на основе метрик RMSE, MAE, MAPE.

Модель	Тренировочная			Тестовая		
	RMSE (\$)	MAE (\$)	MAPE (%)	RMSE (\$)	MAE (\$)	MAPE (%)
LSTM	59.0921	46.9797	0.05465	60.6736	43.1379	0.05224
Fuzzy Gauss	31.6476	22.0033	0.02542	38.1486	23.6430	0.02817

Модель	Тренировочная			Тестовая		
	RMSE (\$)	MAE (\$)	MAPE (%)	RMSE (\$)	MAE (\$)	MAPE (%)
PFS	42.2479	28.1578	0.03246	49.3677	35.7689	0.04172
Fuzzy Activation 1	42.1211	29.6191	0.03453	52.5275	35.3794	0.04288
Fuzzy LSTM	31.6104	16.6874	0.01943	36.7655	18.0011	0.02178

Таблица 3. Сравнение результатов прогнозирования моделей на минутных тиках. Лучшие результаты выделены жирным шрифтом.

Видим, что все предложенные методы фаззификации по отдельности улучшают прогноз. Фаззификация входных данных через гауссианы повышает точность прогноза в 2 раза. А предложенная модель Fuzzy LSTM улучшила прогноз в 2.5 раза.

Рисунок 11 показывает прогнозы модели на основе нечеткой логики на тестовой выборке в сравнении с реальными значениями.



Рис. 11. Сравнение прогноза предложенной модели на основе нечеткой логики и тестовой выборки. Реализация доступна в [9]

6.3.2 Часовые тики

Результаты эксперимента на часовых тиках представлены в таблице 2. Фаззификация по Гауссу в данном случае только ухудшила прогноз, но, зато, параметрическая функция активации отдельно и предложенная модель улучшила прогноз в 1.5 раза.

Модель	Тренировочная			Тестовая		
	RMSE (\$)	MAE (\$)	MAPE (%)	RMSE (\$)	MAE (\$)	MAPE (%)
LSTM	426.6843	283.3210	0.41949	116.9727	72.4595	0.2576
Fuzzy Gauss	429.2130	296.1214	0.4579	184.9171	157.7894	0.5633

Модель	Тренировочная			Тестовая		
	RMSE (\$)	MAE (\$)	MAPE (%)	RMSE (\$)	MAE (\$)	MAPE (%)
Fuzzy Activation 1	400.9681	251.1094	0.36850	127.4488	85.7948	0.30936
Fuzzy LSTM	413.7121	277.0414	0.37167	89.6428	51.9720	0.15919

Таблица 4. Сравнение результатов прогнозирования моделей на часовых тиках. Лучшие результаты выделены жирным шрифтом.

6.3.3 Дневные тики

Результаты эксперимента на дневных тиках представлены в таблице 3. Ситуация аналогична часовым тикам: предложенная модель улучшает прогноз в 2 раза. MAPE улучшилась в 18.5 раз. И фаззификация по Гауссу, и параметрическая сигмоида по отдельности улучшили прогнозы.

Модель	Тренировочная			Тестовая		
	RMSE (\$)	MAE (\$)	MAPE (%)	RMSE (\$)	MAE (\$)	MAPE (%)
LSTM	1189.1016	686.9120	2.78979	269.43616	268.5392	70.5265
Fuzzy Gauss	1296.3405	875.7055	0.54262	669.2094	668.9107	173.163
Fuzzy Activation 1	1185.8253	680.3525	2.6593	189.3476	188.2331	49.522
Fuzzy LSTM	871.0311	392.8022	0.15761	271.1569	234.2518	48.738

Таблица 5. Сравнение результатов прогнозирования моделей на дневных тиках. Лучшие результаты выделены жирным шрифтом.

6.3.4 Время работы

Из-за того, что в моделях увеличилось количество параметров, время обучения нейросети, соответственно, тоже возросло:

Модель	Число дополнительных параметров	Время обучения (сек)		
		1m	1h	1d
LSTM	0	249.614	601.984	53.601
Fuzzy Gauss	$1 \cdot \text{lookback}$	249.472	627.2068	52.727
PFS	$(m - 1) \cdot \text{lookback}$	262.7114		
Fuzzy Activation 1	2	420.686	967.3934	54.6730
Fuzzy LSTM	$2 + 1 \cdot \text{lookback}$	452.313	911.858	56.319

Таблица 6. Сравнение времени обучения моделей.

6.4 Выводы

Предложенная архитектура LSTM на основе нечеткой логики в 2.5 раза увеличивает точность прогноза финансового временного ряда на минутных тиках, в 1.5 раза на часовых и в 2 раза на дневных. Кроме того, на дневных тиках модель уменьшила MAPE в 18.5 раза.

Из-за масштаба данных на часовых и дневных тиках, многие факторы, влияющие на цену в будущем не передаются в нейросеть, поэтому дневной прогноз хуже часового, а часовой, в свою очередь, хуже минутного.

По графикам значений loss-функции от эпохи (Рис. 12), видно, что все модели с первых же эпох достигают асимптоты, поэтому число эпох можно уменьшить до 10 – 20, что уменьшит время обучения.

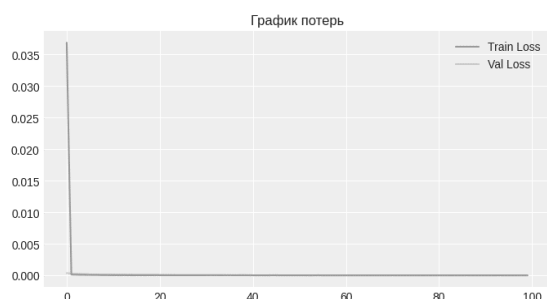


Рис. 12. График потерь для Fuzzy LSTM. Остальные модели ведут себя похожим образом.

7. ЗАКЛЮЧЕНИЕ

В данной работе был разработан вариант архитектуры нейронной сети с памятью на основе нечеткой логики для прогнозирования временных рядов. Введение нечеткости на этапе предобработки данных и внутри модели увеличило точность прогноза в 2.5 раза по сравнению с базовой LSTM на минутных тиках, в 1.5 раза на часовых и в 2 раза на дневных. Такой подход позволил встроить нечеткую логику в процесс обучения модели и повысить интерпретируемость логики работы нейросети.

Полученные результаты открывают новые возможности для применения нечеткой логики и глубокого обучения в задачах прогнозирования финансовых временных рядов.

Перспективными направлениями дальнейших исследований являются:

- Применение альтернативных параметрических функций активации, включая параметризацию \tanh ;

- Использование лингвистических переменных для классификации состояний рынка и тренда;
- Внедрение методов обучения с подкреплением (reinforcement learning) для выработки динамических стратегий прогноза или управления портфелем на основе предсказанных значений.

СПИСОК ЛИТЕРАТУРЫ

- [1] Афанасьева Т. В. Прогнозирование временных рядов: нечеткие модели /Т. В. Афанасьева, А. М. Наместников, И. Г. Перфильева, А. А. Романов, Н. Г. Ярушкина; под науч. ред. Н.Г. Ярушкиной. – Ульяновск: УлГТУ, 2014. – 145 с.
- [2] Andersen T. G. et al. Handbook of Financial Time Series 2009 Springer Berlin, Heidelberg
- [3] Beke A., Kumbasar T. Learning with type-2 fuzzy activation functions to improve the performance of deep neural networks //Engineering applications of artificial intelligence. – 2019. – Т. 85. – С. 372-384.
- [4] Bingham G., Mikkilainen R. Discovering parametric activation functions //Neural Networks. – 2022. – Т. 148. – С. 48-65.
- [5] Гафаров Ф.М., Искусственные нейронные сети и приложения: учеб. пособие / Ф.М. Гафаров, А.Ф. Галимянов. – Казань: Изд-во Казан. ун-та, 2018.
- [6] Hochreiter S. and Schmidhuber J. “Long Short-Term Memory”. In: Neural Computation 9 (Nov. 1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
- [7] Bisht K., Sanjay Kumar, Fuzzy time series forecasting method based on hesitant fuzzy sets, Expert Systems with Applications, Volume 64, 2016
- [8] Кумсков М.И, Арсланова А.Р., Махова А.Г Исследование функций активации и гиперпараметров в архитектуре сетей долгой краткосрочной памяти для прогнозирования финансовых временных рядов. Принято в PRIA
- [9] Makhova A. программная реализация
https://github.com/MakhovaAnastasia/Paper_Fuzzy_LSTM

- [10] Mandal J. K., Bhattacharyya D., Auluck N. Advanced Computing and Communication Technologies //Resource. – 2022.
- [11] Olah C., Understanding LSTM Networks, 2015 <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> (14.11.2024)
- [12] Паклин Н. Нечеткая логика — математические основы <https://loginom.ru/blog/fuzzy-logic>
- [13] Schmidhuber J., Recurrent Neural Networks, 2017 <https://people.idsia.ch/~juergen/rnn.html> (14.11.2024)
- [14] Sezer O. B., Gudelek M. U., Ozbayoglu A. M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019 //Applied soft computing. – 2020. – Т. 90. – С. 106181.
- [15] Ширяев А.Н. Основы стохастической финансовой математики: В 2 т. Т.1: Факты, модели. -М. : МЦНМО, 2016.