



SymbolTemporalMachineT (STMT): ЭФФЕКТИВНОЕ ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ

Аннотация

Современные подходы к прогнозированию временных рядов, особенно основанные на нейронных сетях, сталкиваются с существенными проблемами, такими как переобучение на шумах и катастрофическое забывание. В данной работе мы представляем новый алгоритм, SymbolTemporalMachineT (STMT), разработанный для решения этих проблем. STMT использует символичный подход и механизм контекстной памяти для достижения устойчивости и способности к долгосрочному прогнозированию.

Ключевые слова:

машинное обучение, искусственный интеллект, временные ряды, нейронные
сети

Annotation

Modern approaches to time series forecasting, especially those based on neural networks, face significant challenges such as overfitting to noise and catastrophic forgetting. In this paper, we present a new algorithm, SymbolTemporalMachineT (STMT), designed to address these issues. STMT utilizes a symbolic approach and a context-based memory mechanism to achieve stability and long-term forecasting capabilities.

Keywords:

Machine learning, Artificial intelligence, Time series, Neural networks

1 Введение

Современные алгоритмы прогнозирования временных рядов, в частности подходы на основе нейронных сетей, сталкиваются с рядом фундаментальных проблем. Классические рекуррентные нейронные сети (RNN) и их более сложные вариации, такие как LSTM, испытывают трудности с улавливанием долгосрочных зависимостей [2, 3] из-за проблемы затухания градиентов. Архитектура Трансформер [1] была предложена как решение этой проблемы, используя механизм внимания. На практике Трансформеры не



позволили решить проблему, сталкиваясь с такими трудностями, как нестабильность обучения и большие вычислительные расходы.

В этой статье мы предлагаем архитектуру SymbolTemporalMachineT (STMT) — новую символьную модель, основанную на архитектуре MachineT [4], которая решает эти проблемы. В основе STMT лежит принцип распознавания состояний и переключения контекстных правил, что обеспечивает высокую способность к моделированию сложных циклических процессов. Мы покажем экспериментально, что STMT успешно выполняет долгосрочное прогнозирование в задачах, где нейросетевые модели дают полный провал.

2 Архитектура алгоритма

SymbolTemporalMachineT (STMT) — это алгоритм, основанный на символьном подходе. В его архитектуре каждое дискретное состояние временного ряда представляется одной или несколькими клаузами. Ключевой особенностью STMT является механизм контекстной памяти, который позволяет моделировать сложные последовательности символов, аналогичные предложениям в языке. Как и рекуррентные сети, STMT обрабатывает данные последовательно, но делает это на основе набора выученных логических правил.

2.1 Архитектура алгоритма

В данном разделе мы подробно описываем архитектуру и принципы работы предложенного алгоритма STMT. Сначала мы вводим ключевые определения:

Символ: Базовая единица данных, с которой оперирует модель. Каждый момент времени во временном ряде представляется одним уникальным символом из конечного алфавита $S = \{s_1, s_2, \dots, s_n\}$.

Состояние клаузы: Структура, хранящая степени уверенности в значимости каждого признака и его отрицания.

Клауза: Фундаментальная логическая единица. Каждая из K клауз STMT обучается распознавать определенные темпоральные паттерны. Структурно, клауза C_k включает:



1. **Основную память:** Вектор определяющий правило предсказания символа ассоциированное с данной клаузой.
2. **Контекстную память:** Структура хранящая информацию о типичных последовательностях символов (контекстах), которые предшествовали активации каждой клаузы в прошлом.

Контекстная память: Последовательность из W последних символов, предшествующих текущему моменту времени. Контекст используется моделью для выбора наиболее релевантной клаузы для текущей ситуации.

Предложение: Мы определяем "предложение" как высокоуровневую темпорально-логическую структуру, которая эмерджентно формируется из нескольких клауз. Каждая такая клауза специализируется на определенном участке клаузы в длинной последовательности, а их совместная активация в правильном порядке описывает сложный, долгосрочный паттерн, аналогично тому, как слова складываются в предложение.

Эмерджентность: Мы называем эмерджентностью процесс самоорганизации модели, в ходе которого из взаимодействия простых локальных правил (обновление счетчиков в клаузах) возникают сложные структуры ("предложения"). Эта самоорганизация позволяет модели улавливать фундаментальные закономерности в данных.

2.1.1 Пример работы модели

Представим упрощенную задачу с алфавитом из 5 символов $S = \{A, B, C, D, E\}$ и размером контекстного окна $W=3$.

Предположим, модель уже обработала последовательность и ее текущее Контекстное Окно выглядит так: $[A, B, C]$.

Следующим символом в обучающей выборке является D.

1. **Обучение:** Модель берет символ D и ищет клаузу, которая лучше всего распознает именно его. Допустим, это Клауза №77. Эта клауза является "экспертом по символу D".
 - **Обновление Контекстной Памяти Клаузы №77.** Модель говорит: "Ага, эксперт по D сработал. Какой контекст к этому привел?". Она смотрит в



глобальное Контекстное Окно и видит $[A, B, C]$. Модель укрепляет связь в Контекстной Памяти Клаузы №77 с паттерном $[A, B, C]$. Теперь Клауза №77 "знает", что к появлению D часто приводит последовательность $[A, B, C]$.

- **Обновление Основной Памяти Клаузы №77.** Модель также подкрепляет "уверенность" этой клаузы в том, что она является экспертом по D , еще раз обучая её на символе.

2. Прогноз

- Предположим, в будущем модель видит в своем Контекстном Окне последовательность, очень похожую на $[A, B, C]$.
- Она начинает "опрос" всех клауз, спрашивая: "Для кого из вас этот контекст является знакомым?".
- Клауза №77, которая прочно ассоциировала себя с контекстом $[A, B, C]$, "отвечает" с наибольшей уверенностью.
- Модель выбирает Клаузу №77 как победителя.
- Итог: Поскольку Клауза №77 является "экспертом по символу D ", модель делает вывод, что наиболее вероятный следующий символ — это D .

2.1.2 Обучение

Процесс обучения в STMT происходит для каждого поступающего символа s_t из обучающей последовательности. Цель обучения — настроить основную и контекстную память клауз так, чтобы модель научилась предсказывать следующий символ s_{t+1} на основе контекста. Алгоритм состоит из нескольких шагов:

1. Выбор Активной Клаузы

Для текущего входного символа s_t и предыдущего контекста $C_t = \{s_t - w, \dots, s_{t-1}\}$, модель сначала определяет активную клаузу. Это делается с помощью сравнения текущего контекста C_t с контекстными паттернами хранящимися в контекстной памяти каждой клаузы. Клауза, чей сохраненный контекст соответствует C_t , выбирается как активная клауза C_{active} на данном шаге.

2. Обновление Основной Памяти

После выбора активной клаузы C_{active} её основная память обновляется с целью усилить связь с фактическим следующим символом s_{t+1} . Этот процесс является стохастическим:



- Счетчик соответствующий символу s_{t+1} с вероятностью p увеличивается.
- Счетчики, соответствующие всем остальным символам s_j (где $j \neq t + 1$), с вероятностью p уменьшаются.
- Это заставляет основную память C_{active} “запоминать”, что после паттерна который она распознала должен следовать символ s_{t+1} .

3. Обновление Контекстной Памяти

Одновременно с основной памятью обновляется и контекстная память активной клаузы C_{active} . Модель усиливает ассоциацию между этой клаузой и контекстом C_t , который привел её к активации. Это делается для того, чтобы в будущем, при встрече с похожим контекстом, вероятность выбора именно этой клаузы C_{active} увеличилась.

4. Сдвиг Контекстного Окна:

Наконец, текущий символ s_t добавляется в конец окна контекста, а самый старый символ из него удаляется, подготавливая модель к обработке следующего шага $t + 1$.

2.1.3 Прогнозирование

1. Формирование Входного Контекста

Для предсказания символа в момент времени $t + 1$ модель использует в качестве входа последние W известных символов, формируя входной контекст $C_{in} = \{s_{t-w+1}, \dots, s_t\}$. В режиме долгосрочного и авторегрессионного прогнозирования этот контекст будет состоять из ранее сгенерированных самой моделью символов.

2. Генерация предсказания из Основной Памяти

Выбрав “победившую” клаузу C_{best} , модель обращается к её основной памяти. Этот вектор, по сути, представляет собой выученное правило вида "ЕСЛИ контекст похож на мой, ТО наиболее вероятным будет следующий символ...".

- Значения в основной памяти преобразуются в распределение вероятностей по всем возможным выходным символам. Как правило, положительные значения соответствуют более высокой вероятности, а отрицательные — более низкой.
- Символ с максимальной вероятностью в этом распределении выбирается как итоговый прогноз модели на шаг $t + 1$.

3. Обновление Контекста для Следующего Шага (в режиме авторегрессии)



Если требуется многошаговый прогноз, сгенерированный символ добавляется в конец входного контекста, вытесняя самый старый символ. Этот обновленный контекст будет использован для генерации прогноза на шаге $t + 1$.

3 Сравнение алгоритма с текущими методами

В данном разделе мы проводим экспериментальную оценку предложенной модели STMT. Целью экспериментов является демонстрация производительности и эффективности обучения нашего алгоритма в сравнении с традиционными методами прогнозирования временных рядов.

3.1 Начало эксперимента

Данные:

Для оценки моделей были использованы два набора данных:

1. **Зашумленная Синусоида:** Синтетический временной ряд, сгенерированный функцией синуса с добавлением шума. Этот датасет позволяет оценить способность моделей выделять основной циклический сигнал на фоне случайных помех.
2. **Daily Delhi Climate:** Публичный набор данных, содержащий среднесуточную температуру в Дели. Этот датасет используется для оценки моделей на реальных данных со сложной сезонностью. Обучение проводилось на данных за период с 2013 по 2016 год, а тестирование — на данных за 2017 год.

Основной задачей являлось долгосрочное авторегрессионное прогнозирование. Все модели обучались на тренировочном наборе, после чего генерировали прогноз на всю длину тестового набора, используя на каждом шаге собственные предыдущие предсказания. Качество оценивалось по метрикам R^2 , MAE и RMSE. Для сравнения моделей мы усреднили результаты по трём запускам.

Предложенная модель STMT сравнивалась со следующими базовыми моделями:

- **STMT:** context_window=30, num_bins=150, depth_memory=10, depth_memory_context=3, T=0.
- **AR-Net:** Полносвязная сеть с одним скрытым слоем (64 нейрона, ReLU).
- **RNN/LSTM:** Рекуррентные сети с одним слоем из 32 нейронов.



- **Transformer:** Архитектура на основе внимания с 2 "головами".

Все нейросетевые модели были адаптированы для задачи классификации бинов, используя выходной слой `Dense(num_bins, activation='softmax')`, оптимизатор Adam и функцию потерь `categorical_crossentropy`.

3.2 Результаты экспериментов

Эксперимент 1: Производительность на синтетических данных

На зашумленной синусоиде (Таблица 1) STMT демонстрирует наилучшую производительность, опережая даже архитектуру LSTM. Особенностью модели является скорость обучения: уже после одной эпохи STMT достигает высокого результата ($R^2=0.766$), в то время как все нейросетевые аналоги показывают околонулевые или отрицательные значения, требуя длительного обучения. Это подчеркивает высокую эффективность механизма обучения STMT.

Таблица 1 - Результаты на наборе данных "Зашумленная Синусоида" (после 50 эпох).

Модели для тестирования	R2 (среднее)	MAE (среднее)	RMSE (среднее)
SymbolTemporalMachineT модель (50 эпох)	0.9027	1.0178 °C	2.1578 °C
Авторегрессионная нейронная сеть (50 эпох)	0.8013	2.3084 °C	3.2694 °C
RNN нейронная сеть (50 эпох)	-0.4123	7.2605 °C	8.7394 °C
LSTM нейронная сеть (50 эпох)	0.8971	1.7510 °C	2.3575 °C
Transformer нейронная сеть (50 эпох)	-0.5091	7.4001 °C	9.0308 °C

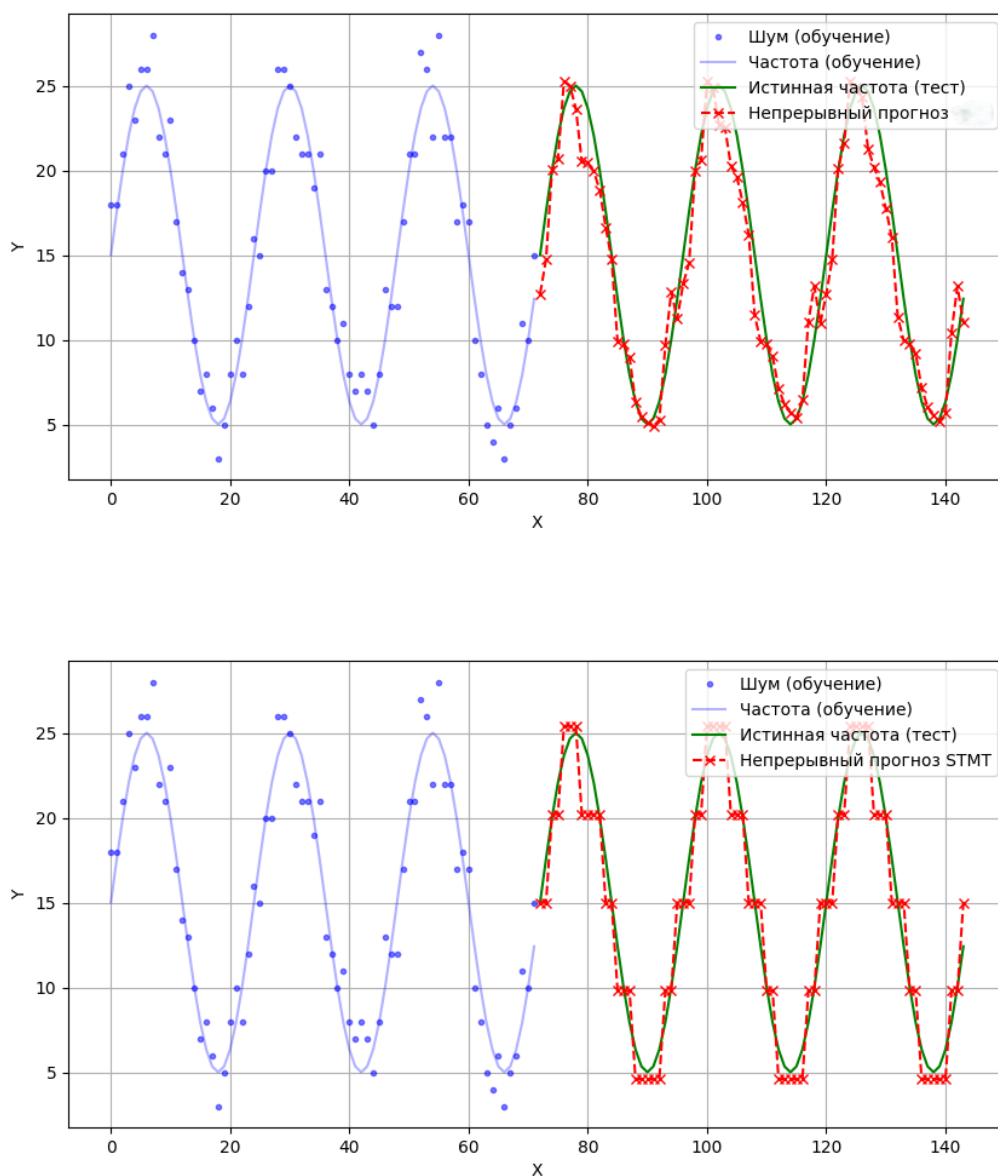


Рисунок 1 - Сравнение прогнозов моделей на синтетических данных. STMT и LSTM.

Эксперимент 2: Производительность на реальных данных

Результаты основного эксперимента на данных о климате в Дели (Таблица 2, Рисунок 2) демонстрируют превосходство STMT.

Таблица 2 - Прогноз среднесуточной температуры в Дейли

Модели для тестирования	R2 (среднее)	MAE (среднее)	RMSE (среднее)
SymbolTemporalMachineT модель (10 эпох)	0.4259	3.8196 °C	4.7979 °C

Авторегрессионная нейронная сеть (10 эпох)	-0.0945	5.6555 °C	6.6165 °C
RNN нейронная сеть (10 эпох)	-0.8090	6.5162 °C	8.0170 °C
LSTM нейронная сеть (10 эпох)	-0.6921	6.2267 °C	8.0764 °C
Transformer нейронная сеть (10 эпох)	-0.4468	5.7944 °C	7.5959 °C

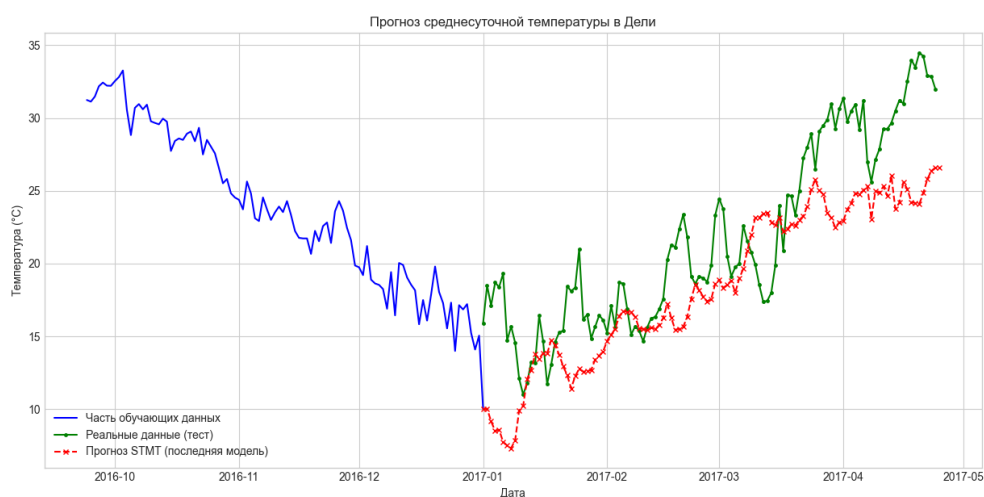


Рисунок 2 - Сравнение прогноза STMT с реальными данными на наборе Daily Delhi Climate.

3.3 Анализ результатов

Полученные результаты демонстрируют, что предложенная модель STMT обладает значительно более высокой эффективностью и стабильностью в задачах долгосрочного авторегрессионного прогнозирования по сравнению с современными нейросетевыми архитектурами.

В то время как модели, LSTM и Transformer, оказались склонными к накоплению ошибок и переобучению на локальных паттернах, STMT успешно справилась с обеими задачами. Мы думаем, что ее успех обусловлен особенностями архитектуры: символьное представление данных фильтрует шум, а ассоциативный механизм поиска по контексту позволяет улавливать глобальную структуру временного ряда, игнорируя локальные аномалии.

4 Выводы

В данной работе мы представили SymbolTemporalMachineT (STMT) — новый символьный алгоритм для обработки временных данных, основанный на механизме контекстной памяти и распознавании состояний. Нашей основной целью было решение проблемы низкой эффективности современных нейросетевых подходов в задачах долгосрочного авторегрессионного анализа.

Проведенные эксперименты на синтетических и реальных данных убедительно продемонстрировали превосходство предложенного подхода. STMT оказалась единственной моделью, способной выполнить стабильный и осмысленный долгосрочный анализ на зашумленных циклических данных, в то время архитектуры LSTM и Transformer, провалились. Мы показали, что успех STMT обусловлен ее структурной устойчивостью к накоплению ошибок, способностью к фильтрации шума через дискретизацию и уникальным механизмом улавливания глобальных паттернов, а не локальных трендов.

Таким образом, основной вклад нашей работы заключается не только в предложении новой, эффективной архитектуры, но и в экспериментальной демонстрации уязвимости современных нейросетевых моделей. Мы утверждаем, что символьные подходы, основанные на интерпретируемых правилах, являются перспективной и более надежной альтернативой "черным ящикам" для целого класса реальных задач.

Продemonстрированная эффективность STMT расширяет область исследований. Мы рассматриваем данную работу как первый шаг к созданию рассуждающих систем, способных к решению сложных целевых задач. Будущие исследования будут направлены на расширение архитектуры для построения иерархических моделей мира, где из простых символов и выученных правил эмерджентно возникают сложные стратегии. Конечной целью является создание прозрачных и энергоэффективных интеллектуальных агентов, которые оперируют внутренней моделью реальности.

Список литературы

1. Han K., Xiao A., Wu E., Guo J., Xu C., Wang Y. Transformer in Transformer // Proceedings of the Neural Information Processing Systems (NeurIPS). 2021. С. 1234–1245. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/854d9fca60b4bd07f9bb215d59ef5561-Paper.pdf (дата обращения: 13.07.2025).



2. Ключев С.Г., Трунов Е.Е. Проблемы обучения глубоких нейронных сетей для обнаружения угроз нарушения безопасности в сетях с динамической топологией // Моделирование, оптимизация и информационные технологии. 2021. Т. 9 (1). С. 1–8. URL: https://web.archive.org/web/20210407072116id_/https://moitvvt.ru/ru/journal/pdf?id=898 (дата обращения: 13.07.2025).
3. Zhang S. D., Tigges C., Biderman S., Raginsky M., Ringer T. Can Transformers Learn to Solve Problems Recursively? 2023. URL: <https://arxiv.org/pdf/2305.14699> (дата обращения: 13.07.2025).
4. MachineT: НОВАЯ АРХИТЕКТУРА ДЛЯ ЭНЕРГОЭФФЕКТИВНОГО МАШИННОГО ОБУЧЕНИЯ. URL: <https://byte-agi.ru/researches> (дата обращения: 13.07.2025).