УДК 004.032.26

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ

О.О. Адиянова[1], К.В. Галанов[1]

[1]Самарский университет, г. Самара
e-mail: ovalga2@gmail.com, galanov.kv@ssau.ru

Аннотация

В статье проведена оптимизация гиперпараметров трёх архитектур нейронных сетей (LSTM, GRU, Transformer) для прогнозирования экономических временных рядов на примере цен на золото. Экспериментально установлено, что модель GRU обеспечивает наилучшую точность прогнозирования при минимальном количестве параметров. Результаты подтверждают эффективность GRU для задач прогнозирования финансовых временных рядов.

Ключевые слова: нейронные сети, GRU, LSTM, Transformer, прогнозирование временных рядов, оптимизация гиперпараметров, цены на золото, MSLE, dropout.

Введение

Прогнозирование финансовых временных рядов - сложная задача из-за высокой волатильности, нелинейности и наличия шумов в данных. Традиционные методы (ARIMA, экспоненциальное сглаживание) часто недостаточно эффективны для учёта долгосрочных зависимостей [1]. Методы глубокого обучения, особенно рекуррентные сети, демонстрируют значительный потенциал.

Наиболее распространены архитектуры LSTM, GRU и Transformer. LSTM эффективно запоминает долгосрочные зависимости [2]. GRU – упрощённая альтернатива LSTM с меньшим количеством параметров [3]. Transformer обрабатывает последовательности параллельно и эффективен на больших объёмах данных [4].

Выбор конфигурации модели критически влияет на точность и устойчивость к переобучению. В литературе недостаточно исследований по сравнению этих архитектур в единых условиях с оптимизацией гиперпараметров.

В работе проводится сравнительное исследование LSTM, GRU и Transformer для прогнозирования цен на золото с систематической оптимизацией: количество слоёв, размер скрытого состояния, dropout, функция активации.

1. Описание эксперимента и модели

В данной работе на одном и том же наборе данных были обучены три модели нейронных сетей: LSTM, GRU и Transformer. Реализация выполнена на языке Python с использованием библиотеки PyTorch. Для оценки точности использовалась метрика MSLE (Mean Squared Logarithmic Error): $MSLE = \frac{1}{n}\sum_{i=1}^{n}(\log(x_i + 1) - \log(\widehat{x_i} + 1))^2$, где x_i- фактическое значение, $\widehat{x_i}$ - прогнозируемое значение, n - количество наблюдений в выборке. Выбор метрики MSLE выбрана, так как чувствительна к относительным

ошибкам и менее чувствительна к крупным выбросам благодаря логарифмической трансформации, что важно для волатильных финансовых рынков.

Все модели обрабатывают последовательности из 15 временных шагов. Данные нормализованы Min-Max Scaling в [0, 1]. Оптимизатор Adam (lr = 0,001), ранняя остановка (patience = 10 эпох).

Датасет содержит ежедневные цены на золото (1979 - 2022 гг.) [5]. Разделение: 80% – обучение, 20% – тестирование (хронологически).

Модель решает задачу одношагового прогнозирования.

2. Результаты оптимизации моделей

Проведена оптимизация гиперпараметров: слои (1-4), hidden_size (16-128), dropout (0,0-0,5). Результаты оптимизации моделей представлены в таблице 1.

Таблица 1 – Лучшие результаты оптимизации моделей LSTM, GRU, Transformer

Модель	**Конфигурация**	**Тест MSLE**	**Val MSLE**	**Эпохи**	**Параметры**
LSTM	layers=2, hidden=32, dropout=0,0	0,184035	0,109159	22	12 961
GRU	layers=2, hidden=32, dropout=0,5	0,026684	0,019005	50	9 729
Transformer	layers=1, d_model=64, nhead=2	0,040990	0,021297	5	844 481

Выводы:

- GRU достигла наименьшей ошибки при минимальном количестве параметров;
- Transformer показал сопоставимую точность, но требует в 87 раз больше параметров;
- LSTM уступил GRU более чем в 6 раз по точности.

На рисунке 1 представлены прогнозы трёх моделей (последние 150 дней). GRU наиболее точно следует реальной динамике, LSTM демонстрирует наибольшие отклонения.

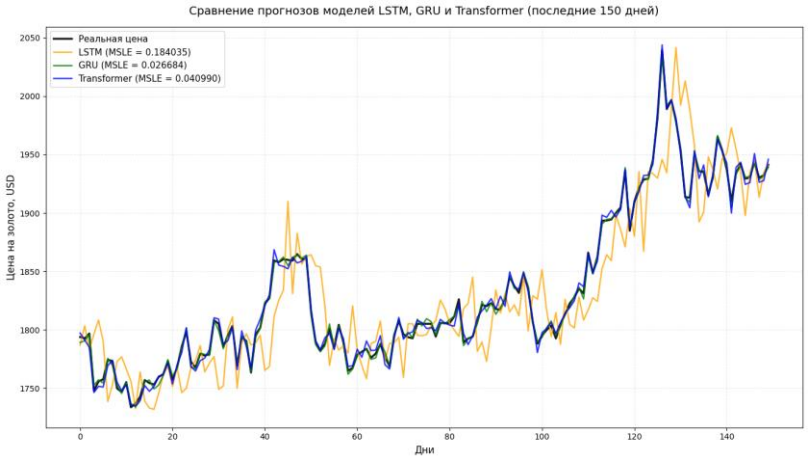


Рисунок 1 – Сравнение прогнозов моделей LSTM, GRU и Transformer (последние 150 дней)

3. Анализ влияния гиперпараметров

В ходе экспериментов выяснилось, что для модели GRU лучше всего использовать два рекуррентных слоя. При 3-4 слоях модель переобучается. Для LSTM - также 2 слоя, для Transformer – 1 слой (добавление слоев вызывает нестабильность, (nan и inf).

При подборе размера скрытого состояния для GRU и LSTM оптимально указывать 32 нейрона, для Transformer – d_model = 64.

Обратим внимание, что в таблице 3 значение тестовой ошибки для hidden_size = 32 формально выше, чем для hidden_size = 16. Результаты получены при фиксированном значении dropout = 0,0. После совместной оптимизации с dropout = 0,5 конфигурация с 32 нейронами показала лучшую обобщающую способность.

Применение dropout оказалось особенно важным для GRU. Для LSTM, напротив, оптимальным оказалось отсутствие dropout. Как видно из таблицы 2, полученной при изолированной настройке параметров, различные гиперпараметры влияют на точность модели GRU.

Таблица 2 – Влияние гиперпараметров на точность GRU
при изолированной настройке параметров

Параметр	Значение	Тест MSLE
Слои	1	0,184219
Слои	2	0,029358
Слои	3	0,179669
hidden_size	16	0,195153
hidden_size	32	0,198931
dropout	0,0	0,213179
dropout	0,5	0,026684

На рисунке 2 показаны кривые обучения GRU для различных dropout. При dropout = 0,0 наблюдается переобучение (валидационная ошибка растёт после 10 эпох). Dropout = 0,2 стабилизирует обучение, но минимальная тестовая ошибка достигнута при dropout = 0,5.

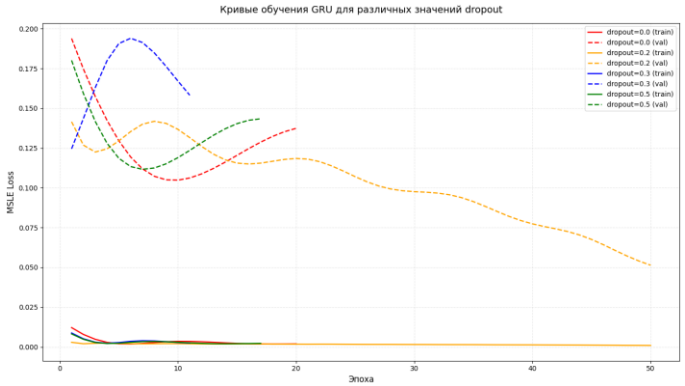


Рисунок 2 – Кривые обучения GRU для различных значений dropout

Заключение

В работе проведено сравнительное исследование архитектур LSTM, GRU и Transformer для прогнозирования цен на золото в единых экспериментальных условиях.

Основные результаты:

1. Модель GRU с конфигурацией (2 слоя, hidden_size = 32, dropout = 0,5) достигла наименьшей ошибки прогнозирования (MSLE = 0,026684) при минимальном количестве параметров (9 729).

2. Transformer показал сопоставимую точность, но требует значительно больше ресурсов.

3. LSTM уступил GRU более чем в 6 раз по точности.

4. Dropout играет ключевую роль в предотвращении переобучения для GRU.

Выбор GRU в качестве предпочтительной архитектуры имеет особое значение для прикладных систем. Меньшее количество параметров снижает требования к памяти и ускоряет выполнение предсказаний, что критично для задач, где решения должны приниматься в реальном времени. Полученные результаты могут быть использованы в системах алгоритмической торговли, риск-менеджмента и финансового анализа.

Список использованных источников

1. Hyndman, R.J. Forecasting: principles and practice / R.J. Hyndman, G. Athanasopoulos. – Melbourne: OTexts, 2021. – 442 p.

2. Hochreiter, S. Long Short-Term Memory / S. Hochreiter, J. Schmidhuber // Neural Computation. – 1997. – V. 9, № 8. – P. 1735–1780.

3. Learning phrase representations using RNN encoder-decoder for statistical machine translation / K. Cho, B. van Merrienboer, C. Gulcehre [et al.] // arXiv:1406.1078. – 2014.

4. Vaswani, A. Attention is all you need / A. Vaswani, N. Shazeer, N. Parmar [et al.] // Advances in Neural Information Processing Systems. – 2017. – V. 30. – P. 5998–6008.

5. Kaggle. Gold Price (1979-Present) Dataset. – URL: https://www.kaggle.com/datasets/ jishnukoliyadan/gold-price-1979-present (дата обращения: 27.03.2025).

OPTIMIZATION OF PARAMETERS OF NEURAL NETWORK MODELS FOR FORECASTING ECONOMIC TIME SERIES

O.O. Adiyanova[1], K.V. Galanov[1]

[1]Samara University, Samara
e-mail: ovalga2@gmail.com, galanov.kv@ssau.ru

Abstract

This article optimizes the hyperparameters of three neural network architectures (LSTM, GRU, and Transformer) for forecasting economic time series using gold prices as an example. It was experimentally established that the GRU model provides the best forecasting accuracy with a minimum number of parameters. The results confirm the effectiveness of GRU for forecasting financial time series.

Keywords: neural networks, GRU, LSTM, Transformer, time series forecasting, hyperparameter optimization, gold prices, MSLE, dropout.