



РАЗРАБОТКА АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ФИНАНСОВЫХ РЫНКОВ: МЕТОДЫ И ОЦЕНКА ЭФФЕКТИВНОСТИ

Иванов Сергей Викторович

научный сотрудник, Белорусский государственный университет
г. Минск, Беларусь

Аннотация

Статья посвящена разработке и анализу алгоритмов машинного обучения, предназначенных для прогнозирования динамики финансовых рынков. В условиях высокой волатильности и стохастичности традиционные эконометрические модели часто оказываются неэффективными. В работе рассматриваются ключевые подходы на основе машинного обучения, способные выявлять сложные, нелинейные зависимости во временных рядах финансовых данных. Особое внимание уделяется методологии, включающей **инженерию признаков** на основе технических индикаторов и сентимент-анализа, а также выбору алгоритмов, таких как **рекуррентные нейронные сети (LSTM)** и **методы градиентного бустинга**. Подробно описывается процесс бэктеста и оценки эффективности модели с использованием таких метрик, как коэффициент Шарпа и максимальная просадка. Анализируются основные вызовы и ограничения, связанные с природой финансовых данных, а также перспективы использования более сложных алгоритмов, например, с применением **обучения с подкреплением**.

Ключевые слова: Машинное обучение, прогнозирование финансовых рынков, временные ряды, нейронные сети, инженерия признаков, бэктестинг, LSTM, алгоритм.

1. Введение

Прогнозирование динамики финансовых рынков всегда было одной из самых сложных и востребованных задач в экономике. Традиционные подходы, основанные на фундаментальном и техническом анализе, часто сталкиваются с ограничениями из-за высокой нелинейности, стохастической природы и «эффективности» рынков, где информация быстро отражается в ценах. Появление и развитие алгоритмов **машинного обучения** открыло новые возможности для решения этой проблемы.

Машинное обучение, в отличие от классических статистических моделей, способно выявлять скрытые закономерности в огромных объемах данных без явного задания правил, что делает его идеальным инструментом для работы с временными рядами. Целью данной статьи является подробное описание процесса разработки, тестирования и оценки алгоритма машинного обучения для прогнозирования финансовых рынков, с акцентом на практические аспекты, связанные с подготовкой данных и выбором оптимальной модели.

2. Особенности данных и инженерия признаков

Финансовые данные представляют собой особый класс временных рядов, которые характеризуются высокой зашумленностью, нелинейностью и нестационарностью. Эти особенности делают их обработку и анализ сложной задачей. Ключевым этапом в разработке успешного алгоритма является **инженерия признаков** — процесс создания новых, информативных переменных на основе исходных данных.

2.1. Исходные данные и их обработка

Для обучения модели используются исторические данные о ценах (Open, High, Low, Close), объемах торгов, а также данные по другим активам. Важной задачей является обработка данных для удаления шума, заполнения пропусков и приведения к стационарному виду, например, через использование логарифмических или процентных доходностей вместо абсолютных цен.

2.2. Создание информативных признаков

Алгоритмы машинного обучения не могут напрямую работать с исходными ценами, поскольку они не содержат достаточной информации для прогнозирования. Поэтому на основе исходных данных создаются новые признаки. Это могут быть:

Технические индикаторы: Скользящие средние, индекс относительной силы (RSI), стохастические осцилляторы, индикаторы волатильности (например, полосы Боллинджера). Эти признаки помогают модели улавливать тренды, перекупленность/перепроданность и импульс.

Признаки, связанные с объемом торгов: Объем торгов часто является индикатором активности и может предвещать движение цен.

Данные сентимент-анализа: Анализ новостных заголовков, постов в социальных сетях и форумов позволяет извлечь информацию о настроениях рынка, которая может быть использована как признак.

3. Основные алгоритмы машинного обучения для прогнозирования

Выбор алгоритма зависит от сложности задачи и характера данных. Для прогнозирования временных рядов наиболее эффективными оказались следующие подходы.

3.1. Алгоритмы на основе деревьев

Random Forest и Gradient Boosting (например, XGBoost или LightGBM) — это ансамблевые методы, которые показывают высокую производительность в задачах классификации и регрессии. Они строят множество решающих деревьев и объединяют их результаты для получения более точного прогноза. Их преимущества заключаются в устойчивости к переобучению, способности работать с нелинейными данными и возможности оценки важности признаков.

3.2. Нейронные сети

Нейронные сети, особенно рекуррентные нейронные сети (RNN) и их модификация LSTM (Long Short-Term Memory), являются одними из самых мощных инструментов для анализа временных рядов. LSTM-сети обладают специальной архитектурой с «воротами», которая позволяет им эффективно обрабатывать последовательные данные и запоминать долгосрочные зависимости, что делает их идеальными для прогнозирования, где важна история временного ряда. LSTM способны «учиться» на последовательности данных, выявляя паттерны, которые предшествуют определенному движению цены.

4. Методология и оценка эффективности

Разработка алгоритма — это только половина дела. Критически важным этапом является его тестирование и оценка эффективности.

4.1. Бэктестинг

Бэктестинг — это процесс тестирования торговой стратегии на исторических данных. Важно проводить его с соблюдением строгих правил, чтобы избежать «утечки данных» (data leakage) и получить реалистичную оценку производительности. Обучение модели должно проходить только на данных, доступных до момента прогнозирования.

4.2. Оценка производительности

Эффективность модели оценивается не только точностью прогноза, но и финансовыми метриками:

Коэффициент Шарпа (Sharpe Ratio): Показывает доходность стратегии на единицу риска. Чем выше коэффициент, тем лучше.

Максимальная просадка (Max Drawdown): Отражает наибольшее падение стоимости портфеля за определенный период. Эта метрика указывает на рискованность стратегии.

Общая доходность и волатильность: Также важные показатели, характеризующие потенциальную прибыль и риск.

Заключение

Разработка алгоритмов машинного обучения для прогнозирования финансовых рынков является сложной, но потенциально очень прибыльной задачей. Успех зависит не только от выбора алгоритма, но и от тщательной подготовки данных, создания информативных признаков и строгой методологии тестирования. LSTM-сети и методы бустинга показывают высокую эффективность в этой области, но ни один алгоритм не может гарантировать 100% точность из-за фундаментальной непредсказуемости рынка. Перспективы развития связаны с использованием более продвинутых методов, таких как **обучение с подкреплением**, которое позволяет модели учиться принимать торговые решения в динамичной среде, а также с интеграцией **квантовых вычислений** для обработки еще больших объемов данных.

Литература

1. Иванов С.В., Петров А.И. (2024). Использование LSTM-сетей для прогнозирования курсов валют. Журнал вычислительных технологий. 25(3), 112–119.
2. Смирнова Е.А., Макаров П.Г. (2023). Инженерия признаков на основе данных сентимент-анализа в задачах прогнозирования. Вестник Белорусского государственного университета. Серия 1: Физика, Математика, Информатика. 12(4), 56–64.
3. Козлов М.В. (2025). Оценка эффективности торговых стратегий на основе машинного обучения: бэктестинг и метрики риска. Финансовая аналитика. 15(1), 78–85.
4. Васильев И.Н., Орлов К.Д. (2024). Применение методов градиентного бустинга для прогнозирования волатильности на фондовом рынке. Инвестиционные технологии. 19(2), 23-30.
5. Новикова А.С. (2023). Обзор архитектур нейронных сетей для анализа финансовых временных рядов. Интеллектуальные системы. 30(4), 45-53.