

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Лотарев Матвей Михайлович, студент 1 курса бакалавриата института экономики и управления АПК, ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, lotarevmat10@gmail.com

Научный руководитель – Анохин Игорь Александрович, ассистент кафедры статистики и кибернетики института экономики и управления АПК, ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, anokhin.igor@rgau-msha.ru

Аннотация. В статье исследуется применение машинного обучения для прогнозирования временных рядов. Автор описывает преимущества и недостатки использования нейронных сетей для этого, в частности, рассматриваются такие подходы как многослойные перцептроны, рекуррентные, сверточные, трансформирующие нейронные сети. Целью статьи является предоставление обзора возможностей применения машинного обучения для прогнозирования временных рядов.

Ключевые слова: машинное обучение, временные ряды, прогнозирование, анализ данных.

APPLICATION OF MACHINE LEARNING TO TIME SERIES FORECASTING

Lotarev Matvey Mikhailovich, student of the 1st year of the Bachelor's degree, Institute of Economics and Management of Agroindustrial Complex, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, lotarevmat10@gmail.com

Scientific supervisor – Anokhin Igor Aleksandrovich, assistant of the Department of Statistics and Cybernetics, Institute of Economics and Management of Agroindustrial Complex, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, anokhin.igor@rgau-msha.ru

Annotation. This paper explores the application of machine learning to time series forecasting. The author describes the advantages and disadvantages of using neural networks for this purpose, in particular, such approaches as multilayer perceptrons, recurrent, convolutional, and transforming neural networks are considered. The purpose of the article is to provide an overview of the possibilities of using machine learning to forecast time series.

Key words: machine learning, time series, forecasting, data analysis.

Временные ряды (ВР) – это последовательности данных, измеренных в определенные моменты времени. Анализ временных рядов является важнейшей задачей во многих областях, таких как экономика, финансы, энергетика, прогнозирование погоды, медицина, транспорт, сельское хозяйство и другие. «Прогнозирование временных рядов является важнейшим аспектом науки о данных, позволяющим компаниям и исследователям прогнозировать будущие значения на основе исторических данных» [4]. Традиционные методы прогнозирования ВР, такие как методы экзогенетического (сезонная декомпозиция, регрессионный анализ и др.) и эндогенного (автокорреляция и др.) прогнозирования, имеют ограничения и не всегда обеспечивают высокую точность прогнозов. Машинное обучение является популярным инструментом для анализа и предсказания будущих значений временных рядов.

Машинное обучение – это раздел искусственного интеллекта, который позволяет компьютерам обучаться на данных и улучшать свои прогнозы без явного программирования. «Подходы машинного обучения (ML) привлекли значительное внимание при прогнозировании временных рядов из-за их способности улавливать сложные закономерности и взаимосвязи в данных» [4], тем самым машинное обучение способно выявлять закономерности в данных, которые могут быть не очевидны для человека, а также может обеспечить более точные прогнозы (особенно для временных рядов с высоким уровнем шума и случайных колебаний), чем традиционные методы, что является неоспоримым преимуществом данного подхода. Ибо это особенно важно для прогнозирования ВР, которые могут иметь сложный и нелинейный характер. Кроме того, машинное обучение позволяет использовать данные в режиме реального времени, что предоставляет возможности по созданию систем прогнозирования, которые могут быстро реагировать на изменения в данных. Машинное обучение также позволяет использовать данные из различных источников для предсказания будущих изменений ВР.

Нейронные сети, способные к обучению на больших объемах данных, выступают как один из наиболее популярных подходов машинного обучения, в том числе для выявления закономерностей временных рядов. Для прогнозирования представляется интересным возможность использования такой их разновидности, как рекуррентные нейронные сети (RNN) или долгая краткосрочная память (LSTM). Эти алгоритмы позволяют выявлять закономерности в данных и прогнозировать будущие значения ВР с высокой точностью. Однако, нейронные сети требуют значительных вычислительных ресурсов и могут быть сложными для интерпретации. Также при применении машинного обучения для предсказания будущих изменений временных рядов необходимо учитывать ряд важных факторов. Например, необходимо выбрать правильную архитектуру нейронной сети, обеспечить качество и полноту данных, используемых для обучения модели. Если пренебречь данными факторами, то это может привести к плохой точности прогнозов.

Методы машинного обучения и как они могут быть использованы для прогнозирования временных рядов представлены в таблице 1.

Таблица 1

Машинное обучение и принципы прогнозирования временных рядов

Подходы машинного обучения	Сущность	Как это работает при прогнозировании временных рядов
Multi-Layer Perceptron (MLP) [Многослойный персептрон]	тип нейронной сети, в которой входные данные проходят через последовательность слоев, включая входной слой, один или несколько скрытых слоев и выходной слой. Каждый нейрон в скрытых и выходных слоях реализует нелинейную функцию, которая комбинирует взвешенные входные сигналы с пороговым значением и преобразует результат	Входное представление ВР в контролируемой задаче обучения реализуется путем создания набора запаздывающих признаков; В процессе обучения нейронная сеть адаптирует свои веса и смещения к закономерностям в обучающих данных путем минимизации функции ошибки прогнозирования. Это достигается с помощью метода обратного распространения; После завершения процесса обучения, MLP использует полученные оптимальные веса для преобразования входных данных в прогнозируемые значения.
Recurrent Neural Networks (RNN) [Рекуррентные нейронные сети]	оснащены механизмом поддержания внутреннего состояния, который позволяет им накапливать и обрабатывать информацию о предыдущих элементах последовательности.	Процесс обработки последовательности осуществляется следующим образом: на каждом временном шаге t сеть получает входные данные x_t и обновляет свое скрытое состояние h_t , используя комбинацию текущих входных данных и предыдущего скрытого состояния h_{t-1} ; Скрытое состояние выступает в качестве внутреннего буфера, который позволяет сети накапливать и передавать информацию о предыдущих элементах последовательности на следующие шаги; Процесс генерации выходных данных включает в себя формирование выходных сигналов на каждом временном шаге, которые могут быть использованы для различных целей, таких как прогнозирование будущих значений или же данные на разных шагах могут быть комбинированы в различных способах.
Convolutional Neural Networks (CNN) [Сверточные нейронные сети]	традиционно используются для обработки данных, но их архитектура может быть адаптирована для прогнозирования ВР путем рассмотрения временного измерения как дополнительной пространственной размерности.	Применяя ядра свертки к входным данным ВР, можно выявлять локальные закономерности и шаблоны в данных; Выборка карт объектов, созданных сверточными слоями, может быть уменьшена для уменьшения размерности и вычислительной сложности; Полностью связанные слои могут объединить извлеченные признаки для получения окончательного результата.

Подходы машинного обучения	Сущность	Как это работает при прогнозировании временных рядов
Decision Tree-Based Models [Модели на основе дерева решений]	модели на основе дерева решений, такие как случайный лес, представляют собой высокоэффективные методы, которые комбинируют несколько деревьев решений для повышения точности прогнозирования и обработки сложных нелинейных взаимосвязей	Данные ВР подвергаются преобразованию в набор функций, которые могут использоваться в качестве входных данных для деревьев решений; Метод случайного леса построен на принципе комбинации нескольких деревьев решений, каждый из которых обучается на части данных и функций, выбранных случайным образом. Затем прогнозы этих деревьев объединяются для получения окончательного прогноза, что позволяет повысить точность.
Transformer Neural Networks [Трансформирующие нейронные сети]	Первоначально разработанный для обработки естественного языка, механизм самоконтроля позволяет модели взвешивать важность различных позиций во входной последовательности, что позволяет ей фиксировать долгосрочные зависимости и взаимосвязи между элементами последовательности.	Архитектура модели позволяет каждой позиции во входной последовательности учитывать всю последовательность в целом, что обеспечивает учет контекста и взаимосвязей между элементами последовательности; Это тип нейронной сети, который состоит из двух основных компонентов: кодировщика и декодера. Кодировщик обрабатывает входную последовательность и преобразует ее в представление. Затем оно используется декодером для генерации выходной последовательности, которая является прогнозом на основе данных.

Источник: Построено автором по данным [4].

Различные исследования показали, что эти методы могут быть эффективными для прогнозирования временных рядов, а также для выявления закономерностей. Так, например, в работе «Прогнозирование временных рядов с применением методов машинного обучения на примере графика выдачи мощности электрической станции» [3] авторы рассматривали подход применения ML к прогнозированию генерации тепловой электрической станции. «Для прогнозирования генерации электрической станции используется модель с тремя последовательными уровнями – на 48 часов вперед и на 24 часа вперед. Каждый уровень (кроме первого) для предсказания использует результат с предыдущих. В качестве результата модель выдает предсказание графика мощности электрической станции с дискретностью 1 час на целевые сутки X» [3]. Авторами статьи были выдвинуты следующие модели: линейная регрессия, LightGBM с экспертными параметрами, LightGBM с подобранными параметрами, CatBoost с подобранными параметрами. Оценки качества созданных авторами моделей, обученных на наиболее значимых признаках, были даны за счет использования метрик качества: MAE, RMSE, MAPE. Было получено, что методы ML показали высокую точность прогнозов, с MAE и MAPE меньше 5%. Авторы заключили, что методы ML могут быть эффективно

использованы для прогнозирования временных рядов в энергетике, что может помочь улучшить управление энергосистемой

В другой работе [5] автор ставит перед собой цель показать эффективность алгоритмов машинного обучения в вопросе прогнозирования цен биржевых торгов. Автором была выбрана биржа СПбМТСБ, а «прогноз средневзвешенных цен реализации ряда нефтепродуктов на 14 дней вперед осуществлялся на основе исходных данных, в качестве которых были предоставлены данные о средневзвешенных ценах торгов по необходимым инструментам, а также некоторые внешние данные, в которых содержится информация об экономических процессах, влияющих на ценообразование» [5]. Автор представил результаты прогнозов моделей ARIMA, VAR, Wavelet и Stacking, а для оценки результатов использовалась метрика MAPE. Исходя из суждений автора можно прийти к заключению, что Модель VAR показала наименьшую точность среди всех рассмотренных моделей, что в свою очередь указывает на то, что она не подходит для прогнозирования цен. Модели Wavelet и ARIMA показали примерно одинаковую точность, что позволяет использовать их в качестве альтернативы друг другу, при этом модель ARIMA может быть использована в качестве базового решения. Более точные результаты продемонстрировала модель Stacking, которая смогла предсказать цену с отклонением менее чем в 2% для большинства рядов. Модель Stacking может быть использована в качестве основы для построения более сложных моделей прогнозирования, которые демонстрируют высокую точность.

Таким образом, в будущем, мы ожидаем, что ML будет продолжать играть важную роль в прогнозировании временных рядов. Однако, необходимо продолжать разработку новых методов и подходов, которые могут улучшить точность прогнозов и сделать ML более доступным и интерпретируемым.

Библиографический список

1. Информационно-аналитическое обеспечение устойчивого развития сельского хозяйства / М. В. Кагирова, В. В. Демичев, Ю. Н. Романцева [и др.]. – Москва : Российский государственный аграрный университет - МСХА им. К.А. Тимирязева, 2023. – 307 с. – ISBN 978-5-9675-2013-6. – EDN RTMIDX.

2. Моделирование уровня производства сельскохозяйственной продукции в крестьянских (фермерских) хозяйствах методами эконометрики и машинного обучения / А. В. Уколова, Б. Ш. Дашиева, Д. В. Быков [и др.] // Вестник Воронежского государственного аграрного университета. – 2023. – Т. 16, № 3(78). – С. 251-262. – DOI 10.53914/issn2071-2243_2023_3_251. – EDN ITNCVJ.

3. Шишков, Е. М. Прогнозирование временных рядов с применением методов машинного обучения на примере графика выдачи мощности электрической станции / Е. М. Шишков, А. В. Проничев, А. А. Савельев // Международный научно-исследовательский журнал. – 2022. – № 2-1(116). – С. 56-60. – DOI 10.23670/IRJ.2022.116.2.009. – EDN GKKBIG.

4. Exploring Machine Learning Approaches for Time Series [Электронный ресурс] / Geeksforgeeks. – Режим доступа: <https://www.geeksforgeeks.org/> (дата обращения 18.10.2024)

5. Прогнозирование цен биржевых торгов на бирже СПбМТСБ [Электронный ресурс] / Spbu – Режим доступа: <https://dspace.spbu.ru/bitstream/> (дата обращения 20.10.2024)

6. Уколова, А. В. Анализ динамики сельскохозяйственного производства хозяйств населения / А. В. Уколова, К. А. Козлов // Экономика и управление: проблемы, решения. – 2024. – Т. 6, № 8(149). – С. 208-220. – DOI 10.36871/ek.ur.p.r.2024.08.06.022. – EDN VBETEQ.

7. Математическая статистика : учебник / А. П. Зинченко, М. В. Кагирова, Ю. Н. Романцева [и др.]. – Москва : Российский государственный аграрный университет - МСХА им. К.А. Тимирязева, 2018. – 199 с. – ISBN 978-5-9675-1681-8. – EDN JIAWIE.

8. Статистико-экономический анализ состояния и развития овцеводства и козоводства в крестьянских (фермерских) хозяйствах Российской Федерации / А. В. Уколова, Б. Ш. Дашиева, В. С. Токарев [и др.] // Известия Тимирязевской сельскохозяйственной академии. – 2023. – № 5. – С. 151-167. – DOI 10.26897/0021-342X-2023-5-151-167. – EDN CXHWCB.