



ИНТЕРПОЛЯЦИЯ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ЭКОСИСТЕМНЫХ ИЗМЕНЕНИЙ

Кузьмина Елена Викторовна

кандидат биологических наук, доцент кафедры биоинформатики, Белорусский государственный университет
г. Минск, Республика Беларусь

Смирнов Алексей Игоревич

магистрант факультета прикладной математики и информатики Белорусский государственный университет
г. Минск, Республика Беларусь

Аннотация

Современные методы машинного обучения (ML) находят всё более широкое применение в экологических исследованиях, в том числе при анализе временных рядов и пространственных данных. В условиях недостаточной наблюдаемости и разреженности экологических измерений интерполяция данных с помощью алгоритмов ML позволяет значительно повысить точность прогнозирования. В статье рассматриваются принципы и методы машинного обучения, применяемые для интерполяции экологических данных, их роль в моделировании и прогнозировании экосистемных изменений. Представлены примеры практического использования нейросетей и деревьев решений в задачах восстановления утраченных данных, а также в моделировании изменения климата, биоразнообразия и водных ресурсов. Обсуждаются преимущества подхода, ограничения и перспективы его развития.

Ключевые слова: машинное обучение, интерполяция данных, прогнозирование, экосистемы, экологическое моделирование, нейросети

Введение

Одной из актуальных задач в области экологии является моделирование динамики природных систем в условиях глобальных изменений климата и антропогенного давления. Однако экологические данные часто носят разрозненный и неполный характер, что обусловлено ограниченностью наблюдательных станций, непредсказуемостью природных процессов и техническими трудностями при проведении замеров. Это особенно актуально при работе с географически распределёнными системами, такими как лесные массивы, водоёмы и пастбища.

В этой связи **методы интерполяции и восстановления пропущенных данных** становятся важным инструментом научного анализа. В последние годы на первый план выходят **алгоритмы машинного обучения**, позволяющие строить высокоточные модели даже при наличии неполной информации. Машинное обучение применяется не только для классификации и предсказания, но и для **интеллектуального заполнения пропусков в данных, пространственной и временной интерполяции**, выявления трендов и аномалий.

1. Машинное обучение как инструмент экологического анализа

Машинное обучение представляет собой раздел искусственного интеллекта, в котором алгоритмы обучаются на основе исторических данных, чтобы делать прогнозы или принимать решения без явного программирования. В экологии ML применяется в следующих задачах:

- **Прогнозирование климатических и гидрологических показателей;**
- **Оценка состояния экосистем и биоразнообразия;**
- **Моделирование распределения видов;**
- **Интерполяция и экстраполяция экологических данных во времени и пространстве.**

Основные алгоритмы, используемые в интерполяции данных, включают:

- **Случайный лес (Random Forest);**
- **Градиентный бустинг;**
- **Многоуровневые нейронные сети;**
- **Методы опорных векторов (SVM);**
- **KNN (метод k ближайших соседей).**

Эти алгоритмы позволяют учитывать нелинейные зависимости, пространственные взаимосвязи и взаимодействие переменных, что особенно ценно для сложных природных систем.

2. Интерполяция экологических данных: подходы и алгоритмы

Интерполяция — это метод оценки неизвестных значений на основе имеющихся данных. В традиционных подходах используют геостатистику (кригинг, инверсия, IDW), однако при больших объёмах разнородной информации и высокоразмерных входных данных они теряют точность. **Машинное обучение превосходит классические методы за счёт способности выявлять сложные скрытые зависимости.**

На практике используются следующие подходы:

- **Пространственно-временная интерполяция:** алгоритмы, такие как RNN и LSTM, анализируют временные ряды с пространственной составляющей,

например, динамику температуры почвы или содержания питательных веществ в воде.

- **Реконструкция данных по спутниковым наблюдениям:** ML-модели обучаются на связке наземных наблюдений и спутниковых снимков, обеспечивая более точную картину распространения видов, состояния лесов или загрязнения.
- **Объединение гетерогенных источников данных:** методы, основанные на ансамблях моделей, интегрируют разные источники (метеоданные, почвенные карты, гидрологические измерения) для восстановления пропущенных значений.

3. Примеры применения в прогнозировании экосистемных изменений

Одним из ярких примеров является использование **глубоких нейронных сетей** для анализа изменений в лесных экосистемах Амазонии. На основе данных о температуре, влажности, выбросах CO₂ и спутниковых изображений были построены модели, прогнозирующие деградацию лесов с высокой точностью.

Другой пример — моделирование изменения уровня подземных вод в засушливых регионах. С помощью **градиентного бустинга** и интерполяции по временным рядам удалось восстановить пропущенные данные о влажности почвы и предсказать дефицит водных ресурсов за 5 лет вперёд.

Также успешно используются **автоэнкодеры** и **генеративные модели**, обучающиеся на исторических климатических данных для генерации альтернативных сценариев изменения экосистем при различных климатических и антропогенных воздействиях.

4. Преимущества, ограничения и перспективы

Машинное обучение значительно повышает точность и надёжность экологических прогнозов, особенно при наличии больших объёмов данных. Среди основных преимуществ:

- Учет нелинейных и комплексных взаимосвязей;
- Высокая устойчивость к шуму и неполноте данных;
- Автоматизация обработки и анализа больших объёмов информации.

Однако существуют и ограничения:

- Необходимость качественных обучающих выборок;
- Риск переобучения и недостаточной интерпретируемости моделей;
- Зависимость от вычислительных ресурсов и точности предобработки данных.

В перспективе ожидается развитие **интерпретируемого машинного обучения (Explainable ML)**, что позволит лучше понять, какие факторы влияют на изменение экосистем. Также важным направлением является **интеграция ML с системной экологией и географическими информационными системами (ГИС)**.

Заключение

Интерполяция экологических данных с применением методов машинного обучения открывает новые горизонты в мониторинге и прогнозировании экосистемных изменений. Это особенно важно в условиях глобального изменения климата, истощения природных ресурсов и необходимости устойчивого управления природной средой. Совмещение экологических знаний с вычислительными возможностями ML позволяет создавать мощные инструменты для научного и практического анализа природы, обеспечивая основу для принятия эффективных решений в области охраны окружающей среды.

Литература

1. Власенко Н. И., Сорока Д. А. Машинное обучение в экологии: принципы и применение. – Минск: БГУ, 2021.
2. Schmidhuber J. Deep Learning in Nature and Ecology // Neural Networks. – 2020.
3. Chen X., Zhang J. Interpolation techniques for geospatial data using machine learning methods // Environmental Modelling & Software. – 2021.
4. Maxmen A. Ecology meets AI: modeling the unpredictable planet // Nature. – 2022.
5. Belgiu M., Drăguț L. Random forest in remote sensing: A review of applications and future directions // ISPRS Journal of Photogrammetry. – 2019.