

Машинное обучение в задачах поддержки принятия решений при управлении охраной природы

К.А. Федутинов

Воронежский государственный университет

Аннотация: Наблюдаемое в последние годы активное развитие методов машинного обучения и анализа данных может повлиять на многочисленные аспекты принятия управленческих решений в экологии. С появлением методов дистанционного зондирования и непрерывного мониторинга большие данные уже вносят свой вклад в таких областях природоохранной деятельности, как прогнозирование погоды, оценка устойчивости окружающей среды и управление стихийными бедствиями. Основной целью данного исследования является выяснение основных концепций применения методов машинного обучения в задачах поддержки принятия решений при управлении охраной природы, а также обсуждение методов повышения эффективности обработки и извлечения знаний из данных экологического мониторинга.

Ключевые слова: экологический мониторинг, нейронная сеть, машинное обучение, теория адаптивного резонанса, ARTMAP, извлечение правил.

Введение

С 2011 года в России функционирует Единая государственная система экологического мониторинга (ЕГСЭМ), введенная Федеральным законом N 7-ФЗ "Об охране окружающей среды". Она состоит из государственной наблюдательной сети, формируемой Росгидрометом, и территориальных систем наблюдений за состоянием окружающей среды, формируемых органами исполнительной власти субъектов. Развитие территориальных систем наблюдений субъектов РФ, техническое оснащение пунктов наблюдений, интеграция всех данных с применением геоинформационных технологий являются стратегическими задачами государства в области охраны окружающей среды.

Общая координация работ по организации и функционированию единой системы мониторинга осуществляется Министерством природных ресурсов и экологии Российской Федерации. На рис. 1. представлена структура взаимосвязей ЕГСЭМ с управляющими структурами регионального и федерального уровня. На региональном уровне данные из

ЕГСЭМ позволяют органам исполнительной власти планировать и обосновывать экологически приемлемые управленческие решения [1].

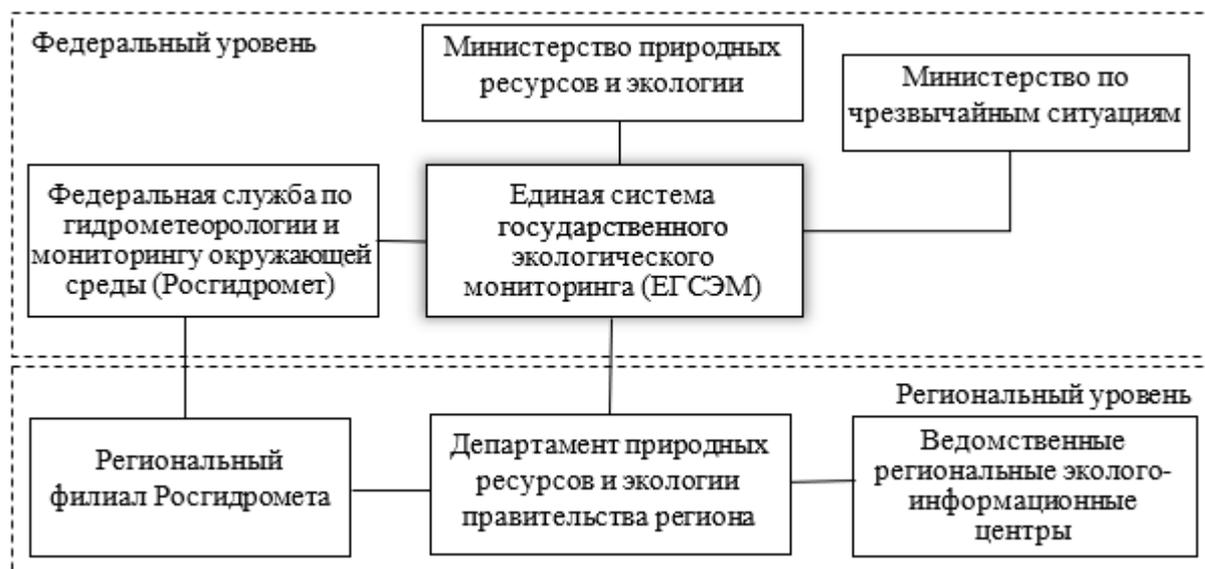


Рис. 1. – Основные связи единой системы государственного экологического мониторинга (ЕГСЭМ)

На рис 2. представлены основные функции региональных органов управления окружающей средой. Первый этап представленной на рис. 2 схемы (накопление данных об объектах управления на основе сетей мониторинга и их структуризации) является ключевым для реализации всех последующих этапов, так как он создает информационную основу для принятия управленческих решений. Для решения задач этого этапа могут использоваться алгоритмы машинного обучения. Интеграция методов машинного обучения с возможностями геоинформационного моделирования позволяет разработать научно обоснованный подход к стратегическому пространственному экологическому управлению регионом и функциональному зонированию территорий. Возрастающие тенденции технологического развития и его влияние на окружающую среду требуют лучшего понимания взаимосвязей между элементами внутри нее [2]. Краткосрочные и долгосрочные изменения в окружающей среде, вызванные

такими изменениями, должны быть наглядно описаны и четко структурированы до проведения любого анализа данных.



Рис. 2. – Основные функции региональных органов управления окружающей средой

Сбор и регистрация больших временных рядов данных и их структуризация для последующего изучения - вот ключевые роли систем экологического мониторинга. Возможность хранить ключевую и достоверную информацию о контролируемых параметрах представляет собой большую проблему, которую необходимо решить. Сбор данных в реальном времени в различных средах открывает новые перспективы в анализе экологических данных. В зависимости от того, используются ли такие данные для лучшего понимания сложившейся ситуации или для прогнозирования и предупреждения о потенциальной опасности, разрабатываются различные методы интеллектуального анализа и поддержки принятия решений [3].

Данные об окружающей среде поступают из самых разных источников, благодаря новым технологиям сбора данных [1]. Большие объемы данных собираются с помощью дистанционного зондирования, когда явления окружающей среды наблюдаются без непосредственного контакта, обычно с помощью спутниковых или бортовых датчиков, включая все более широкое использование дронов. Другие данные собираются с помощью систем мониторинга окружающей среды, которые используют сенсорные технологии в непосредственной близости от наблюдаемых явлений. Такие датчики контролируют ряд параметров атмосферы, литосферы, биосферы, гидросферы и криосферы. Исторически такие сенсорные технологии размещались в полевых условиях и посещались для периодической загрузки данных. Сейчас более распространены телеметрические технологии, обеспечивающие доступ к потокам данных в режиме реального времени. Значительные объемы данных по-прежнему собираются с помощью полевых кампаний, включающих ручное наблюдение и измерение ряда экологических явлений. Примечательно, что при разработке систем управления охраной природы растет интерес к использованию методов сбора данных из Интернета и социальных сетей, таких, как поиск в соцсетях фотографий, показывающих локальные уровни воды в периоды наводнений или, например, поиск сообщений в соцсетях о проблемах с качеством воздуха или воды.

В совокупности перечисленные технологии увеличивает потенциал для получения данных об окружающей среде в беспрецедентном масштабе, что открывает большие возможности для исследований, а также создает ключевые проблемы. В частности, теперь стало очевидно, что исходные данные очень разнородны с точки зрения того, как они должны храниться и извлекаться. Поэтому крайне актуальной становится задача структуризации данных и формирование семантических метаданных, интегрирующих

обширную структурированную информацию о данных и взаимосвязях [2].

Нейросетевые архитектуры для задач обработки экологических данных

Современные модели машинного обучения обеспечивают специализированные вычислительные инструменты для поддержки принятия решений при управлении охраной природы, что позволяет экспертам, занимающимся экологическими данными, оптимизировать возможности для предварительной обработки и структуризации данных крупномасштабного комплексного экологического мониторинга в реальном времени. Нейросетевые алгоритмы превосходят классические алгоритмы машинного обучения по эффективности работы с большими данными благодаря архитектуре параллельной обработки. Рассмотрим примеры нейросетевых структур, которые могут быть полезны для обработки экологических данных (рис. 3).

1. *Сверточные нейронные сети (CNN)*. В CNN данные передаются от входа к выходу, то есть это нейронные сети прямого распространения [3]. Эта архитектура, в основном, используется для обработки изображений, но может хорошо работать со всеми типами данных, включая одномерные данные, такие, как временные ряды показаний датчиков, двумерные данные, такие, как монохромные или цветные изображения, или многомерные данные, такие, как, например, сейсмические данные. Архитектура CNN состоит из сверточных и пулинговых (объединяющих) слоев, за которыми следуют полносвязные слои. Сверточные слои отвечают за извлечение признаков, которое выполняется с использованием операции свертки. Операции свертки используют фильтры (ядра свертки) для формирования карт признаков. Фильтры обрабатывают входные данные, фокусируя внимание на действительно важных входных характеристиках. Объединяющие слои предназначены для уменьшения размерности получаемых карт признаков. Полносвязные слои обрабатывают

результатирующие карты признаков. Преимущество CNN при обработке экологических данных заключается в возможности автоматического формирования признакового пространства, релевантного решаемой задаче.

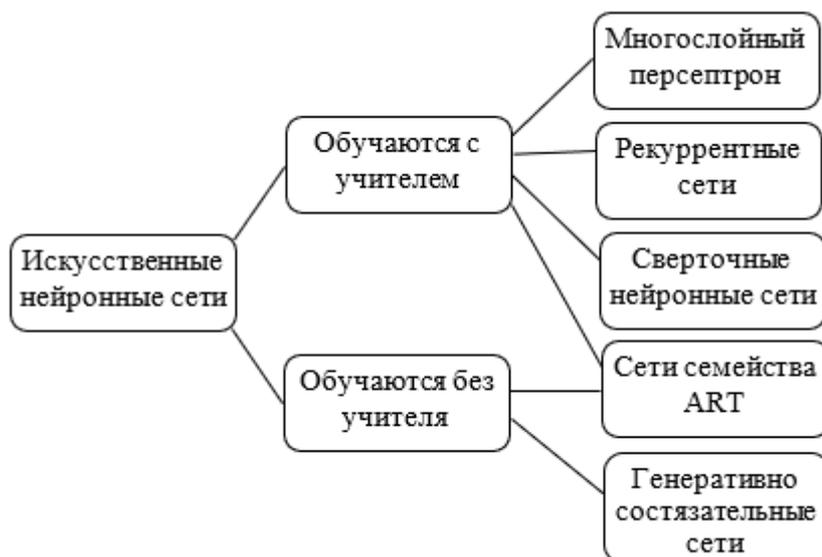


Рис. 3. – Нейросетевые архитектуры для задач обработки экологических данных

2. *Рекуррентные нейронные сети (RNN)*. Наборы экологических данных, как правило, состоят из последовательных временных рядов, содержащих внутреннюю взаимосвязь между образцами данных. Традиционные методы машинного обучения плохо работают с такими последовательными данными. RNN – это специальная архитектура глубокого обучения, которая может моделировать последовательные данные для извлечения закономерностей. В отличие от нейронных сетей прямого распространения, RNN имеет слои с обратной связью. Самый известный пример рекуррентной сети — это сеть с долгой краткосрочной памятью (LSTM) [4]. Модель LSTM – тип рекуррентной нейронной сети, способной обучаться одновременно долгосрочным и краткосрочным зависимостям. Эта архитектура нейронной сети позволяет выделять в мониторинговых данных как долгосрочные тренды, так и настраиваться на текущие изменения, присущие временным рядам измерений, поступающих с различных датчиков.

3. *Многослойный перцептрон (MLP)*. Это одна из простейших (и старейших) архитектур нейронной сети прямого распространения [5]. Входные значения сети MLP передаются на выходной слой, проходя обработку через несколько скрытых слоев. В качестве алгоритма обучения такие сети, как правило, используют базовый алгоритм обратного распространения ошибки. Такого рода сети часто применяются в задачах структуризации и классификации данных.

4. *Генеративно-состязательные сети*. Генеративные состязательные сети (GAN) построены на комбинации двух нейронных сетей: генеративной модели G, которая строит вероятностную аппроксимацию распределения данных в обучающей выборке и на этой основе генерирует новые изображения, и дискриминационной модели D, которая оценивает вероятность того, что выборка получена из обучающих данных, а не сгенерирована генеративной моделью G. Обучение для генеративной модели G заключается в максимизации вероятности ошибки дискриминатора D. Генеративные сети, как правило, используются для генерации реалистичных изображений и хорошо себя проявили при обработке снимков дистанционного зондирования в условиях тумана и плохой видимости [6]. Также они могут принести пользу при обработке данных, содержащих значительное число пропусков.

5. *Сети семейства ART*. Сети семейства ART могут фокусировать внимание на значимых паттернах входных признаков, которые формируются во время обучения сети, подавляя при этом незначимые признаки и шум. Такая особенность позволяет сетям ART успешно преодолевать проблемы, связанные с данными, которые содержат шумы или являются неоднозначно размеченными, а также данными, которые могут быть неверными из-за неисправных датчиков [7]. Важным преимуществом сетей ART является возможность восприятия новых аномальных объектов, без забывания старых.

Современные разновидности сетей ART могут обучаться как "с учителем", так и "без учителя". Базовая архитектура ART - это сеть, которая обучается без учителя. В отличие от большинства нейронных сетей, ART не предусматривает строгого разделения жизненного цикла на стадии обучения и использования. Эти НС обучаются на всех этапах, включая этап эксплуатации, то есть основаны на принципах инкрементного обучения, что наиболее точно отвечает задачам обработки непрерывной поступающей мониторинговой информации. Рассмотрим подробнее возможности использования сетей данной архитектуры для поддержки принятия решений экспертами в природоохранной деятельности.

Применение сетей Cascade ARTMAP в природоохранной деятельности

Адаптивная резонансная теория (ART) является, по сути, когнитивной теорией о том, как наш мозг учится видеть, распознавать и предсказывать объекты и события в быстро меняющемся мире, который может включать в себя множество неожиданностей. Одним из представителей семейства ART является сеть Cascade ARTMAP [8]. Структура такой нейронной сети совместима с представлением знаний в экспертных системах. Это означает, что для экспертов в области управления охраной природы результаты сети могут быть формализованы в виде нечетких следствий "Если a_1, a_2, \dots, a_M , то b_1, b_2, \dots, b_n ", преобразованных в категории распознавания и соответствующих кластерам, сформированных сетью Cascade ARTMAP в результате обработки поступающих на вход сети данных. Правила "Если-то", извлеченные из системы Cascade ARTMAP, аналогичны по форме правилам, получаемым с помощью метода деревьев решений. Однако правила, которые можно извлечь из деревьев решений, и правила Cascade ARTMAP работают совершенно по-разному. Каждое правило дерева решений действует изолированно. Прогноз делается по одному правилу в том случае, если выполняются все его условия. Правила Cascade ARTMAP работают как

ансамбль, управляемый принципом нечеткого выбора, в соответствии с которым каждое правило производит оценку функции выбора сети, чтобы сделать прогноз, и все построенные правила конкурируют друг с другом. На первый взгляд кажется, что легче интерпретировать "точные" правила, такие как в деревьях решений, но на самом деле довольно сложно вырабатывать, скажем, качественные оценки экологического риска, опираясь на точные границы измеренных показателей, особенно, если предположить, что измерения этих показателей могут содержать шумовые и пропущенные значения. Функция нечеткого выбора сети Cascade ARTMAP позволяет частично активировать правило, даже если не все его условия выполнены. Для случаев, когда входы сети попадают в заданные критические диапазоны, функция выбора позволяет полностью активировать правило, но, кроме того, оно может быть частично активировано со степенью уверенности, которая постепенно уменьшается по мере отклонения входов от критических диапазонов. Функция нечеткого выбора сети Cascade ARTMAP обеспечивает нелинейность, необходимую для решения сложноформализуемых задач структуризации разнородной информации [9].

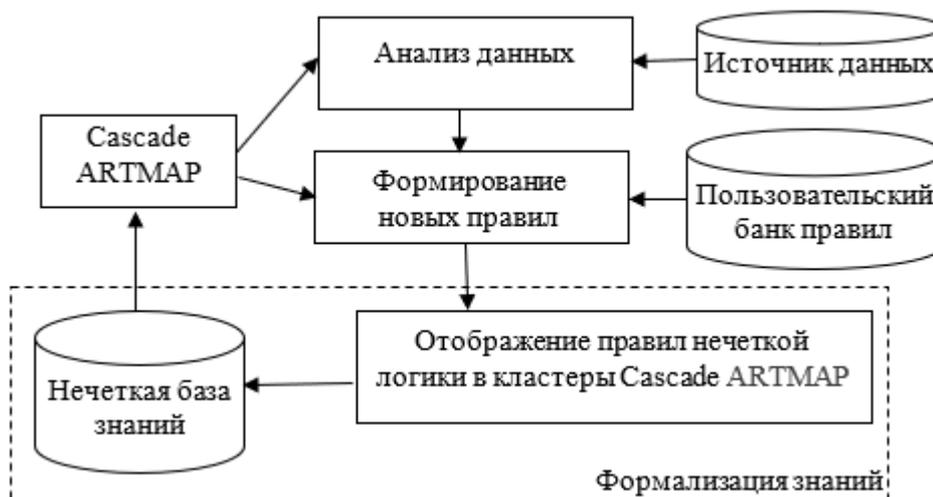


Рис. 4. – Схема обучения сети Cascade ARTMAP

На рис. 4. представлена схема обучения сети Cascade ARTMAP. Информация поступает из источника данных и обрабатывается подсистемой

связывает слой F_2^a с выходным слоем. Ассоциативный паттерн определяется парами векторов $\{A, B\}^p$, ($p=1, 2, \dots$), сформированных в процессе обучения системы. Работу модулей ARTa и ARTb кратко можно описать следующим образом [10]. Входной вектор (соответственно, A или B) после обработки нейронами первого слоя передается на второй слой для отыскания подходящей ему категории (кластера). Вектор прототип найденной категории сравнивается с входным вектором и проверяется, возникло ли состояние резонанса. За состояние резонанса системы отвечает параметр сходства ρ . Если в системе наблюдается резонанс, то входной вектор успешно классифицируется. Если резонанс не наблюдается, это может означать, что система обнаружила новый образец, который не похож ни на один эталонный образец, хранящийся в ассоциативной памяти системы. Выход сети $y(F^{ab})$ после обработки в краткосрочной памяти (STM) может спровоцировать изменение долгосрочной памяти (LTM). Адаптивные веса, хранящиеся в долгосрочной памяти индуцируют нечеткие правила “Если-то”, которые объясняют, какие комбинации признаков и в каком их диапазоне предсказывают те или иные результаты.

Выбор данных, параллельно обрабатываемых модулями ARTa и ARTb, и соединяемых ассоциативными связями включает в себя огромное разнообразие возможностей. Данные, поступающие на вход ARTa, могут включать мониторинговые данные с датчиков, или показатели, используемые для комплексной экологической оценки состояния объектов территориального управления тогда как данные, обрабатываемые ARTb, могут представлять собой показатели здоровья населения или показатели эффективности управления. Как уже отмечалось, адаптивные веса Cascade ARTMAP могут на любом этапе обучения быть переведены в нечеткие правила “Если-то”, которые позволяют экспертам понять природу знаний,

полученных моделью, а также степень вариабельности данных, которую может выдержать каждое из полученных правил.

Заключение

Методы машинного обучения показывают высокую перспективность для оптимизации процесса принятия информационно-ориентированных решений в области охраны природы. Современные интеллектуальные технологии преодолевают разрыв между теоретическими основами методов анализа экологической ситуации и их практической реализацией. Методы машинного обучения помогают принимать обоснованные решения в режиме реального времени, извлекая полезные сведения из больших массивов мониторинговых данных. В частности, основой систем поддержки принятия решений при управлении охраной природы может являться Cascade ARTMAP - самоорганизующаяся нейросетевая система производства и обнаружения правил, с помощью которых в наглядной и понятной форме можно предсказывать изменение окружающей среды.

Литература

1. Lychak, A. I., Vobra T. V. GIS-maintenance of environmental management of territories // Геополитика и экогеодинамика регионов. 2011. Т. 7. № 1-2. С. 126-138.

2. Ефремова О.А. Применение системного подхода к исследованию проблемы использования пространственной информации для поддержки принятия решений региональными органами исполнительной власти // Инженерный вестник Дона, 2014, №2. URL: ivdon.ru/ru/magazine/archive/n2y2014/2371/.

3. Акулов Л.Г., Будко Р.Ю., Вишневецкий В.Ю., Ледяева В.С. Структурное проектирование информационно-измерительных систем для исследования биопотенциалов // Инженерный вестник Дона, 2012, №4. URL: ivdon.ru/ru/magazine/archive/n4y2012/1111/.



ivdon.ru/ru/magazine/archive/n4p2y2012/1476.

4. Ke W., Siyuan Z., Junlan C., Fan R., Lei X. A feature-supervised generative adversarial network for environmental monitoring during hazy days // Science of the Total Environment. 2020. Vol. 748. pp. 141445.

5. Lamba A., Cassey P., Segaran R., Koh L. Deep learning for environmental conservation // Current Biology. 2019. Vol.7. № 29. Pp. 977- 982.

6. Hsieh W. Machine learning in the environmental sciences. Neural networks and kernels. Cambridge: Cambridge University Press. 2009. pp. 274-317.

7. Карпенко С. А. Геотехнологии в управлении территориальным развитием // Ученые записки Таврического национального университета имени В.И. Вернадского. Серия: География. 2010. Т. 23. № 1. С. 149-155.

8. Kashirina I.L., Fedutinov K.A., Bondarenko Y.V. Development of information and analytical tools based on adaptive classifier Cascade ARTMAP// Journal of Physics: Conference Series 2020. Vol. 1479. p. 12066.

9. Федутинов К.А. Структуризация экологической информации с применением геоинформационных технологий // Моделирование, оптимизация и информационные технологии. 2019. Т. 7. № 4. С. 131-140.

10. Каширина И.Л., Федутинов К.А. Кластеризация непрерывного потока данных на основе обобщенной модели нейронной сети семейства ART // Системы управления и информационные технологии. 2018. № 1. С. 33-39.

References

1. Lychak, A. I., Bobra T. V. Geopolitika i ekogeodinamika regionov 2011. Vol. 7. No. 1-2. pp. 126-138.

2. Efremova O. A. Inzhenernyj vestnik Dona, 2014, № 2. URL: ivdon.ru/ru/magazine/archive/n2y2014/2371

3. Akulov L. G., Budko R. Yu., Vishnevetsky V. Yu., Ledyeva V. S. Inzhenernyj vestnik Dona, 2012, № 4. URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1476



4. Ke W., Siyuan Z., Junlan C., Fan R., Lei X. Science of the Total Environment. 2020. Vol. 748. P. 141445
5. Lamba A., Cassey P., Segaran R., Koh L. Current Biology. 2019. Vol.7. № 29. Pp. 977- 982.
6. Hsieh W. Machine learning in the environmental sciences. Neural networks and kernels. Cambridge: Cambridge University Press. 2009. pp. 274-317.
7. Karpenko S. A. Uchenye zapiski Tavricheskogo nacional'nogo universiteta imeni V.I. Vernadskogo. Seriya: Geografiya. 2010. Vol. 23 (62). № 1. pp. 149-155.
8. Kashirina I. L., Fedutinov K. A., Bondarenko Y. V. Journal of Physics: Conference Series 2020. Vol. 1479. P. 12066.
9. Fedutinov K. A. Modelirovanie, optimizaciya i informacionnye tekhnologii. 2019. Vol. 7. № 4. pp. 131-140.
10. Kashirina I. L., Fedutinov K. A. Sistemy upravleniya i informacionnye tekhnologii.2018. Vol. 71. № 1. Pp. 33-39.