

## Геосистемный подход в развитии методов и алгоритмов анализа пространственных данных

*С.А. Ямашкин, А.А. Ямашкин, О.А. Зарубин, А.В. Кирюшин*

*МГУ им. Н.П. Огарёва, Саранск*

**Аннотация:** В статье представлено исследование, направленное на разработку методов и алгоритмов анализа пространственных данных для диагностики состояния геосистем. Показано, что объединение моделей машинного обучения в ансамбль позволяет повысить устойчивость анализирующей системы: точность решений, принимаемых ансамблем, имеет тенденцию стремиться к точности наиболее эффективного моноклассификатора системы. Расчет и консолидация территориальных дескрипторов при этом позволяют снизить размерность анализируемых данных, облегчить допустимую емкость модели машинного обучения, повысить ее устойчивость к переобучению, не допустить значительного снижения точности классификации в рамках конкретной решаемой задачи.

**Ключевые слова:** метагеосистемы, пространственные данные, тестовые полигоны, территориальные дескрипторы, ансамбли.

### Введение

Прогресс в области технологий цифрового картографирования и анализа пространственных данных и материалов дистанционного зондирования Земли, развитие методического и алгоритмического обеспечения процесса анализа структуры земель обусловили повышение спроса на географическую информацию [1]. Инфраструктуры пространственных данных (ИПД) обеспечивают использование географической информации в решении экономических, социальных и экологических проблем на местном, региональном и национальном уровнях.

Если рассматривать анализируемые территории, как «геосистемы», – «особый класс управляющих систем; земное пространство всех размерностей, где отдельные компоненты природы находятся в системной связи друг с другом и как определенная целостность взаимодействуют с космической сферой и человеческим обществом» [2, с. 292], то важнейшей задачей наук о Земле становится исследование многообразия взаимосвязей, существующих между природой и обществом. Данная научная проблема в настоящее время решается при условии тесного междисциплинарного

взаимодействия с учеными и специалистами в области технических наук и информационных технологий, предполагает целенаправленное использование достижений в области интеллектуального анализа данных.

Целью данной работы является характеристика результатов в области разработки новых методов и алгоритмов анализа пространственных данных на основе отечественного учения о геосистемах для принятия управленческих решений в области устойчивого развития природно-социально-производственных систем.

Решение научной проблемы разработки методов и алгоритмов интеграции, обработки, интеллектуального анализа и распространения пространственных данных актуально в соответствии с опорными положениями Стратегии научно-технологического развития Российской Федерации (утверждена Указом Президента Российской Федерации от 1 декабря 2016 г. № 642) для достижения возможности перехода к передовым цифровым технологиям в области устойчивого пространственного развития и территориального планирования.

### **Методология и методы исследования**

По данным международных реферативных баз данных и систем цитирования Scopus и Web of Science по состоянию на 2022 год, характерен устойчивый рост числа статей, написанных в рамках междисциплинарных исследований в области разработки новых эффективных алгоритмов анализа пространственных данных.

В Постановлении Правительства Российской Федерации от 1 декабря 2021 года №2148 об утверждении государственной программы Российской Федерации «Национальная система пространственных данных» отмечено, что для обеспечения совершенствования инфраструктуры пространственных данных (ИПД) требуется реализация мероприятий, направленных на решение проблем «... ограниченного использования современных российских

геоинформационных технологий и высокопроизводительной обработки пространственных данных». О высокой научной значимости и актуальности решения обозначенной проблемы говорит позитивный опыт других стран. Директива 2007/2/ЕС Европейского парламента и Совета от 14 марта 2007 г. о создании инфраструктуры пространственной информации в Европейском сообществе (INSPIRE) направлена на последовательное внедрение ИПД государствами-членами Европейского Союза.

Решение задачи анализа пространственных данных возможно на основе использования технологий глубокого машинного обучения, позволяющих достичь высокой точности интерпретации пространственных данных за счет анализа нелинейных зависимостей [3, 4]. Не будет преувеличением отметить и то, что при внедрении моделей и алгоритмов интеллектуального анализа данных, задача формирования исходных информационных массивов практически всегда представляет собой более сложный и ответственный процесс в сравнении с последующим этапом обучения и тонкой настройки моделей [5]. Если пространственные данные для анализа содержат в себе достаточное количество информации для выявления определенных закономерностей и гипотез, то автоматизированный анализ имеет все шансы быть успешным. Напротив, какими бы эффективными и сложными ни были применяемые интеллектуальные алгоритмы, они не дадут достичь необходимого результата при недостаточной полноте и репрезентативности анализируемых данных.

Ответ на вопрос о том, как повысить эффективность алгоритмов машинного обучения при анализе пространственных данных, можно найти в отечественных учениях о геосистемах [6], в рамках которых пространственно-временная структура территориальных систем может быть представлена в виде процессов обмена и трансформации энергии, вещества и информации.

---

В соответствии с геосистемным подходом, вмещающая геосистема оказывает существенное влияние на свойства иерархически подчиненных образований. Полезным источником информации о ней являются данные дистанционного зондирования Земли. Пространственные данные, характеризующие определенную территорию, должны быть определены как совокупностью признаков локального объекта (базовый уровень геопространственной модели), так и знаниями о вмещающих иерархических геосистемах. При проведении исследований используется типологическая классификация ландшафтов по В. А. Николаеву [7], предусматривается следующая иерархия таксономических единиц: разряд, подразряд, класс, подкласс, тип, подтип, род, подрод ландшафтов.

Наконец, реальную практическую ценность процесс анализа пространственных данных приобретает при достижении возможности распространения полученных результатов исследований или показателей мониторинга территориальных систем. Эффективное решение этой задачи возможно с помощью проектирования, внедрения и использования проектно-ориентированных геопортальных систем, функционирующих на основе современных веб-технологий и позволяющих предоставить доступ к актуальным пространственным данным на основе технологий облачного хранения [8]. Информативная визуализация пространственных данных о геосистемах осуществляется на основе цифровых карт, функционирующих в рамках геопортальных веб-интерфейсов.

Целесообразно выделить два направления развития каркаса геопортальной системы. С одной стороны, необходимо придерживаться дедуктивной стратегии, при которой геопортал изначально разрабатывается с целью достижения возможности покрытия решений максимального количества проектных задач. С другой – необходимо максимально эффективно реализовывать индуктивную стратегию, в рамках которой

---

положительно зарекомендовавшие себя улучшения частных конкретных геопорталов становятся основой для развития каркаса системы. Совместное использование дедуктивной и индуктивной стратегий обеспечит эволюционное улучшение каркаса построения геопортальных систем, а также оптимизацию частных решений, разрабатываемых на его основе.

### **Новые модели и алгоритмы анализа пространственных данных**

Анализ иерархической структуры геосистем показывает, что разработка методов и алгоритмов анализа и интеграции информации в инфраструктурах пространственных данных может быть раскрыта через систему тестовых полигонов. С целью разработки и апробации новых методов и алгоритмов анализа пространственных данных, создана система тестовых полигонов, раскрывающих особенности взаимодействия лесостепных и лесных геосистем Приволжской возвышенности и Окско-Донской низменности.

Методы и алгоритмы машинного обучения могут быть эффективно использованы в целях интерпретации геопространственных данных, для которых характерны свойства пространственной зависимости, пространственной неоднородности, масштабируемости и соподчиненности [9]. Нейросетевые модели, на вход которых подаются данные, как о конкретной территории, так и о вмещающих геосистемах, способны показывать более высокую точность классификации.

Интересный результат дает опыт построения ансамблей классификаторов, предполагающий необходимость объединения нейросетевых моделей в ансамбль для усиления устойчивости анализирующей системы [10]. Так, точность решений, принимаемых ансамблем, имеет тенденцию стремиться к точности наиболее эффективного моноклассификатора системы, а ошибочность системы имеет тенденцию не

превышать ошибочность наиболее эффективного классификатора, избегая систематических ошибок, допускаемых отдельными классификаторами.

Использование ансамблей позволяет подойти к решению проблемы подготовки данных для обучения моделей за счет интеграции в систему моделей, обученных на различных комбинациях обучающих и валидационных выборок для снижения влияния ошибок, возникающих при формировании датасетов. Построение эффективных ансамблей может быть основано на моделях относительно небольшой ширины и глубины, что позволяет проектировать высокоточные классификаторы, обучение которых менее требовательно к вычислительным мощностям в сравнении с классическими глубокими моделями.

Другой важный результат исследований заключается в разработке методики расчета территориальных дескрипторов, системный анализ которых позволяет достичь прироста точности классификации геосистем [11]. Учет предложенной системы дескрипторов, рассчитанных на основе данных дистанционного зондирования Земли (ДЗЗ) (в том числе, метрик ландшафтного разнообразия), цифровой модели рельефа и ландшафтной карты, позволил достичь точности, не уступающей сверточным сетям. Учет дескрипторов рельефа увеличивает точность на 3 %, а ландшафтных метрик – на 11 %. Картограммы представленных дескрипторов хорошо интерпретируются специалистами в области наук о Земле.

### **Проектная реализация геопортальных систем**

Геопортальная система «Природное и культурное наследие Республики Мордовия» представляет собой пример проектной реализации геопортальных технологии и выполняет функцию организации управления системами культурного и исторического наследия. Каркас проекта изначально проектировался и разрабатывался на основе компонентно-ориентированного подхода для эффективного решения задачи распространения информации о

природном и культурном наследии региона. Графические интерфейсы разработаны с использованием технологий адаптивной верстки для достижения UX-показателей и удобства использования со смартфонов и десктопных устройств (рис. 1).

Информация в геопортале представлена на 3 языках: русском, английском и испанском. Данные в рамках проекта структурированы в 69 информационных блоках из 6 тематических разделов: «Административное устройство, природа и ресурсы», «История», «Народонаселение и социальная инфраструктура», «Экономика», «Экология», «Наследие муниципальных районов» и снабжены иллюстративным материалом: более 800 фотографий объектов природного, исторического и культурного наследия. Основу контентного наполнения формируют более 1240 информационных блоков о населенных пунктах с информацией о топонимике, географии, истории размещены на геопортале с привязкой к электронной карте, 830 объектов культурно-исторического наследия и 95 особо охраняемых природных территорий описаны с пространственной привязкой к цифровой карте региона.

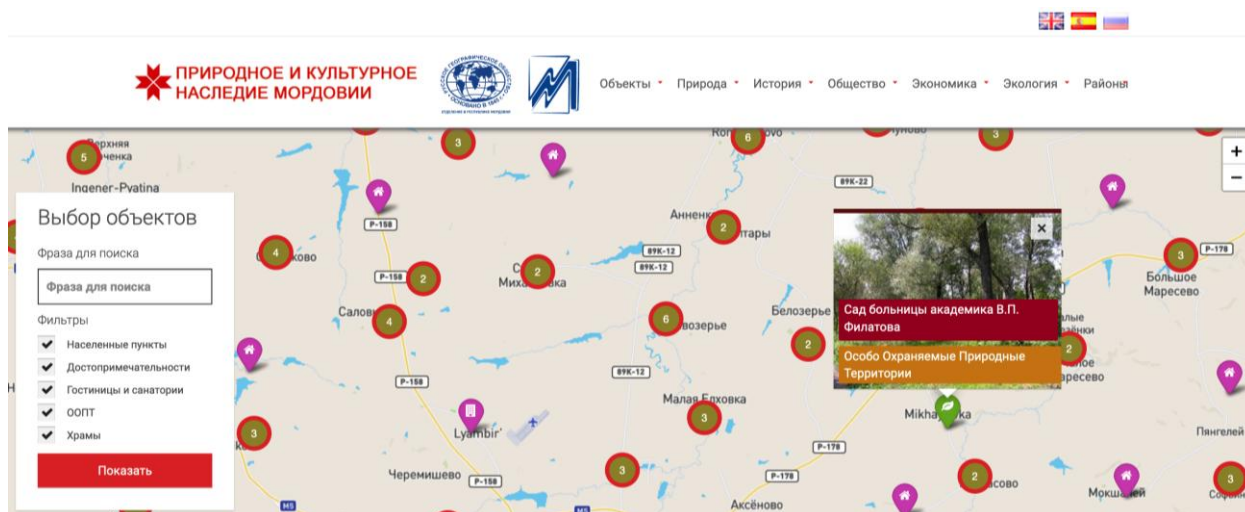


Рис. 1. – Цифровая карта геопортала «Природное и культурное наследие Республики Мордовия»

На основе материалов региональной ИПД, синтезированных с применением новых методов и алгоритмов машинного анализа пространственных данных, в рамках проекта разработано 139 тематических масштабируемых карт различной тематической направленности. Карты растительности, ландшафтов и эрозионных процессов разработаны и актуализируются на основе новых методов и алгоритмов.

### **Выводы**

Развитие машинного обучения в последние годы – это период, когда результаты российских исследований в области ландшафтоведения смогут проявить себя на острие научно-технологического развития. Геосистемный подход позволяет осмысленно подойти к решению задачи подготовки данных для автоматизированного анализа и повысить точность моделей машинного обучения.

Использование ансамблей, выстраиваемых по предложенной методике, обеспечивает возможность проводить оперативный автоматизированный анализ пространственных данных для решения задачи тематического картографирования метагеосистем и природных процессов. Объединение моделей в ансамбль позволяет повысить устойчивость анализирующей системы: точность решений, принимаемых ансамблем, имеет тенденцию стремиться к точности наиболее эффективного моноклассификатора системы. Ошибочность системы в большинстве случаев не превышает ошибочность наиболее эффективного классификатора, при этом исключается вероятность грубых систематических ошибок, допускаемых отдельными классификаторами. Формирование метаклассификатора по предложенному алгоритму представляет собой возможность добавить элемент предсказуемости и контроля в использование нейронных сетей, традиционно представляющих собой «черный ящик». Анализ дескрипторов даёт возможность снизить размерность анализируемых данных, но в то же время





приводит к утрате определенного количества информации. Поиск баланса позволяет облегчить емкость модели машинного обучения и повысить ее устойчивость к переобучению.

Практическую ценность интегрируемые в ИПД массивы пространственных данных приобретают при условии наличия возможности их гибкого использования представителями определенных ролей при принятии взвешенных управленческих решений. Построение проблемно-ориентированных геоинформационных систем, обеспечивающих процесс междисциплинарного исследования метагеосистем для решения задачи управления территориальными системами, должно опираться на фундаментальные принципы и методы, определяющие научно обоснованные способы анализа больших массивов пространственных данных, синтеза новых знаний о природно-социально-производственных системах, построения программно-аппаратных систем хранения, визуализации и распространения пространственно-временных данных посредством современных веб-технологий.

### **Благодарности**

*Исследование выполнено за счет гранта Российского научного фонда № 22-27-00651, <https://rscf.ru/project/22-27-00651/>.*

### **Литература**

1. Касимов Н. С., Мазуров Ю. Л., Тикунов В. С. Концепция устойчивого развития: восприятие в России // Вестник Российской академии наук. 2004. Т. 74. № 1. С. 28-36.
2. Сочава В. Б. Введение в учение о геосистемах. – Новосибирск : Издательство "Наука" Сибирское отделение, 1978. 320 с.
3. Галушка В.В., Фатхи В.А. Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз

данных // Инженерный вестник Дона. 2013. № 2. URL: ivdon.ru/magazine/archive/n2y2013/1597.

4. Лиля В. Б. Алгоритм и программная реализация адаптивного метода обучения искусственных нейронных сетей // Инженерный вестник Дона. 2012. № 1. URL: ivdon.ru/magazine/archive/n1y2012/626.

5. LeCun Y., Bengio Y., Hinton G. Deep Learning // Nature. 2015. № 521. 436-444.

6. Yamashkin A. A., Yamashkin S.A., Muchkaeva N. S., Lyamzina I. S. Development of the regional water balance regulation concept based on the geosystem approach // Indonesian Journal of Electrical Engineering and Computer Science. 2022. Vol. 26. № 3. P. 1672-1683.

7. Николаев В.А. Классификация и мелкомасштабное картографирование ландшафтов. – М.: Изд-во Московского университета, 1978. – 62 с.

8. Кошкарев А. В., Тикунов В. С., Тимонин С. А. Картографические веб-сервисы геопорталов: технологические решения и опыт реализации // Пространственные данные: в информационных, кадастровых и геоинформационных системах. 2009. № 3. С. 6-12.

9. Черкашин А. К. Теоретическая и метатеоретическая география // Географический вестник. 2020. № 1(52). С. 7-21.

10. Ямашкин А.А., Ямашкин С.А. Анализ метагеосистем полигона «Инерка» посредством ансамблей моделей машинного обучения // ИнтерКарто. ИнтерГИС. – 2022. – 28 (Ч1). – С. 613–628.

11. Ямашкин С. А., Ямашкин А.А. Классификация метагеосистем с применением моделей машинного обучения // Геодезия и картография. – 2022. – Т. 83. – № 7. – С. 25-38.

### References

1. Kasimov N. S., Mazurov Ju. L., Tikunov V. S. Vestnik Rossijskoj akademii nauk. 2004. T. 74. № 1. pp. 28-36.



2. Sochava V. B. Vvedenie v uchenie o geosistemah [Introduction to the study of geosystems]. Moskva: Izdatel'stvo "Nauka" Sibirskoe otdelenie, 1978. 320 p.
3. Galushka V.V., Fathi V.A. Inzhenernyj vestnik Dona. 2013. № 2. URL: [ivdon.ru/magazine/archive/n2y2013/1597](http://ivdon.ru/magazine/archive/n2y2013/1597).
4. Lila V. B. Inzhenernyj vestnik Dona. 2012. №. URL: [ivdon.ru/magazine/archive/n1y2012/626](http://ivdon.ru/magazine/archive/n1y2012/626).
5. LeCun Y., Bengio Y., Hinton G. Nature. 2015. № 521. pp. 436-444.
6. Yamashkin A. A., Yamashkin S.A., Muchkaeva N. S., Lyamzina I. S. Indonesian Journal of Electrical Engineering and Computer Science. 2022. Vol. 26. № 3. pp. 1672-1683.
7. Nikolaev V.A. Klassifikacija i melkomasshtabnoe kartografirovanie landshaftov [Classification and small-scale landscape mapping]. Moskva: Izd-vo Moskovskogo universiteta, 1978. 62 p.
8. Koshkarev A. V., Tikunov V. S., Timonin S. A. Prostranstvennyye dannye: v informacionnyh, kadastryh i geoinformacionnyh sistemah. 2009. № 3. pp. 6-12.
9. Cherkashin A. K. Geograficheskij vestnik. 2020. № 1(52). pp. 7-21.
10. Yamashkin A.A., Yamashkin S.A. InterCarto. InterGIS. 2022. 28 (P1). pp. 613-628.
11. Yamashkin S. A., Yamashkin A.A. Geodezija i kartografija. 2022. vol. 83. № 7. pp. 25-38.