

На правах рукописи

Колесников Алексей Александрович



Разработка методологии использования
искусственного интеллекта в цифровой картографии

1.6.20. Геоинформатика, картография

Автореферат диссертации на соискание ученой степени доктора
технических наук

Новосибирск – 2025

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Сибирский государственный университет геосистем и технологий» (СГУГиТ).

Научный консультант – доктор технических наук, профессор Лисицкий Дмитрий Витальевич.

Официальные оппоненты:

Матерухин Андрей Викторович, доктор технических наук, федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет геодезии и картографии», декан факультета геоинформатики и информационной безопасности;

Присяжнюк Сергей Прокофьевич, доктор технических наук, профессор, Закрытое акционерное общество «Институт телекоммуникаций», генеральный директор;

Ступин Владимир Павлович, доктор технических наук, доцент, федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский национальный исследовательский технический университет», профессор кафедры маркшейдерского дела и геодезии.

Ведущая организация – Федеральное государственное автономное образовательное учреждение высшего образования «Пермский государственный национальный исследовательский университет» (г. Пермь).

Защита состоится 29 апреля 2025 г. в 13-00 часов на заседании диссертационного совета 24.2.402.02 при ФГБОУ ВО «Сибирский государственный университет геосистем и технологий» по адресу: 630108, Новосибирск, ул. Плахотного, 10, ауд. 402.

С диссертацией можно ознакомиться в библиотеке и на сайте ФГБОУ ВО «Сибирский государственный университет геосистем и технологий»:

<https://sgugit.ru/science-and-innovations/dissertation-councils/dissertations/kolesnikov-aleksey-aleksandrovich/>

Автореферат разослан 14 февраля 2025 г.

Ученый секретарь

диссертационного совета



Дубровский Алексей Викторович

Изд. лиц. ЛР № 020461 от 04.03.1997.

Подписано в печать 24.01.2025. Формат 60 × 84 1/16.

Печ. л. 1,0. Тираж 100 экз. Заказ 6.

Редакционно-издательский отдел СГУГиТ
630108, Новосибирск, Плахотного, 10.

Отпечатано в картопечатной лаборатории СГУГиТ
630108, Новосибирск, Плахотного, 8.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Необходимость автоматизации картографических процессов обусловлена существенными изменениями окружающего мира и соответствующими новыми требованиями к геопространственной индустрии, в том числе автоматизации, высокой динамике получения и передачи геоинформации и знаний во времени, близком к реальному. Новые информационные технологии, разнообразие источников и скорость формирования новых данных требуют эволюции в геопространственной инфраструктуре посредством автоматизации при создании картографических произведений и геоинформационных моделей в экономике, обществе и окружающей среде.

Объемы и способы использования пространственных данных в том или ином виде постоянно увеличиваются, становятся более разнообразными, что приводит к совершенствованию существующих и появлению новых технологий, методик, алгоритмов для их обработки, анализа и использования, позволяет решать задачи эффективнее, но одновременно влечет за собой сложность однозначного выбора в том множестве знаний, которые можно использовать для целей автоматизации процессов цифровой картографии. Дополнительным фактором становится то, что для работы с пространственными данными современный уровень развития компьютерных средств и технологий позволяет использовать технологии искусственного интеллекта, методологическая часть которых значительно отстает от объемов и глубины практического использования.

С точки зрения направлений стратегии мирового научного развития, тема исследования соответствует национальной стратегии развития технологий искусственного интеллекта, принятой в рамках XXIII Петербургского международного экономического форума, является составной частью сквозных технологий национальной программы «Цифровая экономика Российской Федерации», соответствует дорожной карте развития технологий искусственного интеллекта Российской Федерации, отвечает требованиям разрабатываемого стандарта ISO/IEC «Artificial intelligence. Concepts and terminology» и Указу Президента РФ от

10.10.2019 № 490 (ред. от 15.02.2024) «О развитии искусственного интеллекта в Российской Федерации».

На современном уровне развития геоинформационных технологий, в условиях постоянного значительного увеличения разнообразия, сложности, комплексности и актуальности пространственных данных, постоянного расширения доступных для использования методов и средств обработки пространственных данных, чрезвычайно большого объема программных средств, который недостаточно структурирован и обобщен, существует проблема отсутствия систематизации и фильтрации накопленных научных и практических знаний в области обработки пространственных данных, теории и методологии использования систем искусственного интеллекта для автоматизации решения задач цифровой картографии. Это имеет место на всех этапах создания и использования картографических произведений и геоинформационных моделей.

Говоря о цифровой картографии, нужно отметить, что сама сущность и принципы обработки пространственных данных с помощью инструментов автоматизации (в том числе и на основе технологий искусственного интеллекта) имеет ряд уникальных особенностей:

– вариативность исходных данных. Это обуславливается тем, что объекты карты имеют семантику с указанием времени (требуется алгоритмы и методы обработки табличных данных и временных рядов), для наибольшей актуальности картографических материалов используются снимки со спутников и беспилотных летательных систем (БАС) (алгоритмы и методы компьютерного зрения, а также обработки временных рядов), и часто дополнительную информацию извлекают из социальных сетей, лент новостей, отчетов (алгоритмы и методы обработки естественного языка). Также необходимо учитывать, что во многих исследованиях исходные геопространственные данные являются четырехмерными, то есть представляют собой временные ряды, у каждого измерения которых существует пространственная привязка (дополнительно требуются алгоритмы и методы обработки многомерных разновременных данных). Алгоритмы

анализа и прогнозирования временных рядов и пространственных отношений достаточно хорошо проработаны по отдельности, но их сочетание относительно мало описано и исследовано;

– разнообразие представления координатной и атрибутивной составляющих пространственных данных; разные алгоритмы и методы обработки ориентированы на тот или иной вид входных данных. Это могут быть пары или тройки координат, форматы файлов или баз данных геоинформационных систем, а также производные варианты (инжиниринг данных) на основе пространственных характеристик объектов и их взаимоположения: сводные характеристики объекта (площадь, периметр, количество узлов и т. д.); нахождение в определенной области (например, с использованием различных регулярных сеток); характеристики относительно других объектов (расстояния до ближайшего, количество объектов в радиусе и т. д.) и комбинации вышеперечисленного. При этом, опять же в зависимости от используемых алгоритмов и методов, производные параметры нужно формировать определенным образом. Но в случае с пространственными данными очень редко в описании алгоритмов и методов имеются рекомендации по формированию наиболее оптимальных способов представления;

– большое количество существующих алгоритмов обработки всех видов данных (как имеющих координатную привязку, так и нет); при переходе к обработке данных в задачах исследования пространственных данных часто возникает необходимость выбора наиболее подходящего алгоритма (или нескольких, для последующего сравнения) для решаемой задачи. Блок-схемы выбора технологий искусственного интеллекта в зависимости от задачи и существующих данных предлагаются как исследователями в области обработки данных, так и разработчиками программного обеспечения соответствующего направления, но их недостатками с точки зрения картографии и геоинформатики является практически полное отсутствие таких критериев, как наличие пространственной составляющей в данных и их разброс по времени, а это достаточно сильно влияет на последующее качество построенной математической модели. Поэтому важным будет

предложить блок-схему выбора технологии искусственного интеллекта, построенную на основе уже существующих блок-схем и дополненную условиями о пространственных характеристиках объекта и изменяемости его параметров с течением времени. В качестве дополнительных критериев представляется актуальным добавить параметры интерпретируемости математической модели и точность полученных результатов с точки зрения выбранного способа оценки. Интерпретируемость результатов и построенной математической модели является достаточно важной для оценки корректности результата и его дальнейшего использования. Поэтому различные виды линейной регрессии и деревьев решений остаются популярными в противовес нейронным сетям, результаты которых, как правило, значительно лучше, но представляют собой «черный ящик». Говоря о точности при выборе алгоритма, нужно учитывать, какая именно метрика (способ оценки точности результата) взята для оценки работы алгоритма. С точки зрения пространственного анализа, для проверки корректности предсказаний положения объекта в пространстве и, особенно, времени применяется несколько вариантов, зависящих от решаемой задачи, исходных данных, требований к объяснимости результатов;

– различные подходы к решению типовых задач; современная цифровая картография включает в себя очень разнородные по процессам и данным задачи и направления деятельности, которые также могут быть с той или иной степенью автоматизированы с помощью технологий искусственного интеллекта: навигационные системы, web-картография, 3D-картография, геостатистика, цифровые двойники, тактильная картография и ассистивные ГИС, восприятие геовизуализации, ГИС оперативного мониторинга, пространственные вопросно-ответные системы, оценка и прогнозирование на основе геометрических и атрибутивных данных, классификация и сегментация изображений.

Перечисленные направления и задачи могут быть объединены общими методами их решения и комплексным картографо-геоинформационным подходом, реализованным с помощью современных информационных и геоинформацион-

ных технологий, математических алгоритмов и технических решений. При этом во всех этих задачах могут быть использованы все вышеперечисленные элементы, причем в разных комбинациях. То есть данные могут быть обработаны одним алгоритмом, затем выполнен инжиниринг данных и затем еще один этап обработки другим алгоритмом. Также в ряде задач можно обрабатывать данные несколькими алгоритмами параллельно для уточнения итогового результата (ансамблирование алгоритмов). И все эти особенности также рассмотрены только для отдельных конкретных задач, без общих рекомендаций по построению подобных конвейеров обработки (англ. pipelines). При этом из-за очень большого числа комбинаций метод перебора для поиска наиболее оптимального варианта невозможен.

В соответствии с перечисленными особенностями и сложностями обработки пространственных данных в рамках диссертационной работы разработаны методология и теория использования систем и технологий искусственного интеллекта, которые позволяют сформулировать рекомендации и методики обработки данных в зависимости от потребностей, задач и процессов цифровой картографии, что повышает степень автоматизации и роботизации процессов, существенно сокращает затраты времени и, в ряде случаев, повышает точность обработки данных.

На современном уровне развития геоинформационных технологий, в условиях постоянного значительного увеличения разнообразия, сложности, комплексности и актуальности пространственных данных (в том числе и в неявном варианте), постоянного расширения доступных для использования методов и средств обработки пространственных данных, чрезвычайно большого объема программных средств, который недостаточно систематизирован, структурирован и обобщен, существует проблема отсутствия систематизации накопленных научных и практических знаний в области обработки пространственных данных, теории и методологии использования систем искусственного интеллекта для автоматизации решения задач цифровой картографии. Это имеет место на всех эта-

пах создания и использования картографических произведений и геоинформационных моделей. Например, при автоматизированном дешифрировании данных дистанционного зондирования (спутниковых снимков, облаков точек) в программном обеспечении обычно присутствует несколько методов (алгоритмов), основанных на принципах искусственного интеллекта, но при этом недостаточно или полностью отсутствуют указания, какой именно метод (алгоритм) следует применять для каких задач и исходных данных. То же самое справедливо и для ряда других картографических задач, например, извлечения географических названий из неструктурированного текста, расчета прогнозных значений и т. д.

Систематизация научных и практических знаний с использованием новой методологической и технологической информационной основы использования систем искусственного интеллекта для обеспечения автоматизации картографического и геоинформационного производства с помощью научно-обоснованного выбора методов и средств до выполнения предлагаемых исследований ранее не решалась.

Чрезвычайно большой объем информации о данной предметной области слабо систематизирован и обобщен, недостаточно структурирован. Существующие разработки в области оптимального выбора методов и средств искусственного интеллекта для обеспечения картографической деятельности имеют разрозненный характер, недостаточно соответствуют принципам системного подхода, не опираются на единую методологию и общность технологической реализации в картографии с учетом современного научно-технического уровня. В настоящее время развитие методов искусственного интеллекта, разработки технических решений для обработки, преобразования и анализа данных различных типов, появление аппаратных средств с гораздо большими вычислительными ресурсами открывают принципиально новые возможности для автоматизации и роботизации картографических процессов. Исследования по данной теме позволят оптимально использовать методы и технологии искусственного интеллекта для решения задач большинства существующих технологий обработки пространственных данных

и, в конечном итоге, повысить общий технологический уровень и оперативность цифровой картографии.

Степень разработанности темы. Автоматизация и роботизация остается одним из приоритетных направлений в картографии и геоинформатике, особенно в последнее время с учетом новых потребностей экономики, социальной сферы и обороны, и новых возможностей развития программных и аппаратных средств, алгоритмов, методов и технологий. Средства и методы автоматизации технологических процессов картографии и геоинформатики разрабатывались и ранее, но это выполнялось большей частью на основе написания компьютерных программ, которые полностью проектировались специалистами и выполняли заранее указанные действия по обработке пространственных данных, без использования тех возможностей, которые предоставляют технологии искусственного интеллекта. То есть был задан порядок и условия выполнения шагов, значения параметров и т. п., в отличие от принципа искусственного интеллекта, когда компьютерной программе предлагаются только исходные данные и то, как именно они должны быть обработаны в итоге, а промежуточные шаги формируются автоматически на основе поставленной пользователем задачи. Значительный вклад в развитие теории и методов картографии и геоинформатики внесли отечественные и мировые ученые в части общей теории геоинформационного картографирования – Берлянт А. М., Тикунов В. С., Tobler W., в части формализации картографических процессов – Васмут А. С., в части автоматизации процессов создания тематических карт Верещака Т. В., Ширяев Е. Г., в части геоинформационного моделирования – Журкин И. Г., Иванов А. Г., Истомин Е. П., Карпик А. П., Присяжнюк С. П., Сербенюк С. Н., Флоринский И. В., Шихов А. Н., Buckley A., в части интеллектуализации технологических процессов – Лисицкий Д. В., Лурье И. К., Майоров А. А., Мартыненко А. И., Поспелов Д. А., Розенберг И. Н., Савиных В. П., Цветков В. Я., Schmidt A. H., в области обработки пространственно-временных данных – Бешенцев А. Н., Братков В. В., Матерухин А. В., Погорелов А. В., в части использования данных дистанцион-

ного зондирования – Долгополов Д. В., Ступин В. П., Черных Е. Г., Шаповалов Д. А., Sen Gupta A. K.

Выполненные исследования и разработки в области автоматизации картографических и геоинформационных процессов упростили и оптимизировали обработку и преобразование больших объемов пространственных данных, проверку качества цифровых карт. Однако вышеуказанная проблемная ситуация полностью не решена из-за наличия в картографии плохо математически обусловленных процессов, основанных на интеллектуальных способностях человека и трудно поддающихся автоматизации. В настоящее время развитие методов искусственного интеллекта, разработки технических решений для обработки, преобразования и анализа данных различных типов, появление аппаратных средств с гораздо большими вычислительными возможностями открывают принципиально новые возможности для автоматизации картографических процессов.

Исследования по данной теме позволят оптимально использовать методы и технологии искусственного интеллекта для решения задач автоматизации большинства существующих технологических процессов и, в конечном итоге, повысить общий технологический уровень цифровой картографии.

Цель исследования – разработка теории и методологии использования искусственного интеллекта для автоматизации решения задач цифровой картографии.

Для достижения этой цели были поставлены следующие *задачи исследования*:

1 выполнить анализ объектов исследования с целью оценки текущего состояния и выявления проблем использования технологий искусственного интеллекта для решения задач цифровой картографии;

2 исследовать существующие алгоритмы, методы и технические решения, используемых в технологиях искусственного интеллекта и систематизировать существующие варианты, способы и ограничения их использования применительно к задачам цифровой картографии;

3 разработать принципы, теоретические основы и понятийный аппарат, позволяющие обеспечить автоматизацию и роботизацию процессов картографирования с использованием технологий искусственного интеллекта;

4 разработать методологию использования технологий искусственного интеллекта, включающую в себя теоретически обоснованный логический аппарат выбора и разработки: способа представления пространственных данных, алгоритмов и методов, критериев оценки качества их работы, оптимального конвейера обработки данных при решении определенной задачи цифровой картографии, базирующихся на технологиях искусственного интеллекта;

5 реализовать необходимые технические решения, позволяющие использовать технологии искусственного интеллекта в геоинформационных системах и сервисах при выполнении процессов цифрового картографирования.

Объект и предмет исследования. Объектом исследования являются принципы, теория и методология использования технологий искусственного интеллекта для автоматизации и роботизации решения задач цифровой картографии, предметом исследования является цифровая картография.

Решаемая проблема или разрабатываемая научная гипотеза. Постоянный рост объемов пространственных данных приводит к тому, что традиционными методами они не могут быть обработаны с достаточной для потребителей скоростью и качеством. Данная проблема решается использованием технологий искусственного интеллекта для решения задач цифровой картографии путем систематизации накопленных научных и практических знаний в области обработки пространственных данных, создания теории и методологии и разработки на этой основе технических решений для автоматизации и роботизации технологических процессов. Кроме этого, существующее множество алгоритмов, методов и технологий искусственного интеллекта, которые потенциально могут быть применены при автоматизации и роботизации процессов цифровой картографии, недостаточно проработаны и требуют доработки и актуализации своды формализо-

ванных правил и рекомендаций по их применению и интеграции в традиционные процессы в зависимости от решаемой задачи цифровой картографии.

Научная новизна исследований автора заключается в следующем.

1 Разработана концепция, систематизирующая теоретические положения и технические решения по автоматизированному построению конвейеров обработки пространственных данных, создаваемых на основе элементов программного обеспечения, использующего технологии искусственного интеллекта.

2 Разработаны новые технологические схемы, использующие алгоритмы, методы и технические решения искусственного интеллекта, ускоряющие автоматизированное создание картографических и геоинформационных моделей по разнородным данным (данные дистанционного зондирования, фотографии, текст и т. д.), в том числе и с учетом временной составляющей.

3 Разработан логический аппарат и теоретические основы выбора критерия качества обработки пространственных данных при решении определенной задачи цифровой картографии с использованием технологий искусственного интеллекта, а также предложены новые способы их вычисления для отдельных процессов.

4 Разработана новая методология подбора оптимальной последовательности операций обработки пространственных данных на основе технологий искусственного интеллекта с учетом анализа их свойств и возможностей относительно автоматизации и роботизации решаемых задач для ускорения выполнения типовых процессов цифровой картографии.

Перечисленные разработки реализуют предлагаемую методологию использования технологий искусственного интеллекта для решения задач цифровой картографии и формируют на ее основе новую технологию автоматизации и роботизации единого комплексного цикла создания, редактирования и использования карт с использованием алгоритмов и методов искусственного интеллекта, которая позволит кратно увеличить скорость и функциональность на всех этапах – от сбора информации до выпуска картографических произведений.

Теоретическая значимость работы заключается в развитии теории и методологии цифровой картографии на основе исследований возможностей и путей использования технологий искусственного интеллекта для увеличения степени автоматизации картографических процессов. В рамках данной работы рассматриваются следующие элементы этой теории: понятийный аппарат, логический аппарат выбора наиболее оптимальных шагов создания и редактирования картографических произведений, оценки качества, классификации технологий и алгоритмов искусственного интеллекта с точки зрения обработки пространственных данных, система критериев оценки применимости технологий искусственного интеллекта для целей цифровой картографии.

Практическая значимость:

1 Разработана база данных алгоритмов искусственного интеллекта, способов формирования синтетических (инжиниринга) данных, методик оптимизации сформированных математических моделей и подбора гиперпараметров, методов оценки точности при обработке пространственных данных, предназначенная для экспертной системы, работающей на основе семантических сетей для подбора наиболее оптимального конвейера обработки пространственных данных алгоритмами, методами и техническими решениями искусственного интеллекта.

2 Выполнена формализация использования технологий искусственного интеллекта для автоматизации обработки всех типов пространственных данных, построения и обновления картографических произведений с использованием баз знаний и больших языковых моделей.

3 Создано наукоемкое информационное и программное обеспечение для распространенных геоинформационных систем, позволяющее использовать технологические элементы работы для автоматизации процессов создания и редактирования картографических произведений, а также коллективных многоплановых исследований технологий искусственного интеллекта для задач обработки пространственных данных.

4 Разработана методика автоматизированной стилизации и оформления картографических материалов на основе существующих картографических произведений.

5 Разработаны рекомендации по интерпретации математических моделей процессов, создаваемых в рамках решения задач цифровой картографии, получаемых на основе пространственных данных с использованием технологий искусственного интеллекта.

Методология и методы исследования основаны на системном анализе, математическом моделировании объектов и процессов, теории математической обработки результатов измерений. Решение поставленных задач базируется на использовании математического моделирования, системного анализа и синтеза, сравнения, обобщений и оценок с привлечением методов интегрального исчисления, численных методов и теории интерпретации полученных результатов, а также геоинформационного анализа и структуризации данных.

Предложенная методология позволяет использовать методы и технологии искусственного интеллекта для решения задач цифровой картографии, оптимизировать процессы выбора алгоритмов технологических процессов, автоматизировать процессы создания и редактирования картографических произведений и сервисов. Элементы разработанной методологии применения технологий искусственного интеллекта являются основой для формирования тематических карт различного содержания, например прогнозирования количества случаев заболеваний тропическими болезнями по картографическим и климатическим данным, обновления данных картографических сервисов на основе сегментации спутниковых снимков, извлечения географических названий и построение карт на основе неструктурированного текста, уточнения данных дистанционного зондирования Земли (ДЗЗ) для построения карт типов растительности по картографическим данным, прогнозирования стоимости объекта недвижимости по картографическим данным, текстовому описанию и фотографиям, rapid mapping территорий после чрезвычайных ситуаций, семантической сегментации облаков то-

чек, комплексного картографирования и моделирования объектов гидроэнергетики, сельского хозяйства и другой тематики.

Положения, выносимые на защиту:

1 Предлагаемые методологические принципы использования технологий искусственного интеллекта позволяют автоматизировать и значительно ускорить решение геоинформационных и картографических задач, вплоть до обеспечения режима реального времени;

2 Предлагаемые методологические принципы автоматизированного формирования конвейеров геоинформационной и картографической обработки данных на основе семантических баз данных, больших языковых моделей и мультимодальных нейронных сетей позволяют устранить проблему между наличием широкого спектра методологических, технологических, алгоритмических и программных средств искусственного интеллекта, обладающих существенно разными особенностями, возможностями, ограничениями, характеристиками и недостаточностью обоснованных указаний по их использованию при решении определенных задач цифровой картографии;

3 Предлагаемая формализация способов оценки качества и интерпретации математических моделей обработки пространственных данных, получаемых на основе технологий искусственного интеллекта, позволяет повысить качество конечных геоинформационных продуктов – цифровых геоинформационных моделей, цифровых карт, баз пространственных данных;

4 Разработанная методология обработки пространственных данных с помощью технологий искусственного интеллекта, основанная на построении семантической структуры, обеспечивает автоматизированный выбор элементов (методов, алгоритмов) и построение технологических цепочек (конвейеров обработки данных) из существующих программных инструментов создания и использования цифровых карт и геоинформационных моделей;

5 Апробация разработанных технологических и программных решений и методов их качественной оценки позволила определить эффективность пред-

лагаемой методологии использования технологий искусственного интеллекта для автоматизации решения задач цифровой картографии.

Соответствие диссертации паспорту научной специальности. Диссертационное исследование по содержанию и характеру полученных результатов соответствует следующим областям исследования: 6 – Технические средства и технологии сбора, хранения и обработки пространственных и пространственно-временных данных. Оперативный анализ и картографирование потоковой географической информации. Геосенсорные сети и датчики, 11 – Геоинформационные системы (ГИС). Математическое, информационное, лингвистическое и программное обеспечение ГИС и их приложений, 19 – Большие данные в задачах геоинформационного и картографического моделирования. Разнородные, разномасштабные и разновременные пространственные данные, вопросы их интеграции и совместного использования. Применение искусственного интеллекта для обработки пространственных данных паспорта научной специальности 1.6.20. Геоинформатика, картография, разработанного экспертным советом ВАК Минобрнауки России.

Степень достоверности и апробация результатов исследования.

Основные результаты исследований, концептуально-теоретические положения диссертации докладывались и обсуждались на следующих научно-технических конференциях: Международном научном конгрессе «Интерэкспо ГЕО-Сибирь» в период с 2017 по 2024 г., (г. Новосибирск); Национальной конференции «Регулирование земельно-имущественных отношений в России: правовое и геопространственное обеспечение, оценка недвижимости, экология, технологические решения» в период с 2018 по 2020 г. (г. Новосибирск); Международной конференции «ИнтерКарто. ИнтерГИС» (19–21 июля 2018 г., г. Петрозаводск; 2020, 2022, 2024 гг., онлайн); Всероссийской конференции «Преподавание информационных технологий в Российской Федерации» (2018 г., г. Москва); Международной конференции «Региональные проблемы дистанционного зондирования Земли» (2018 г., г. Красноярск; 2020 г., онлайн); Международной конферен-

ции «Gi4DM 2019 GeoInformation for Disaster Management» (3–6 сентября 2019 г., Чехия, г. Прага); Всероссийской конференции с международным участием «Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM)» (26–29 августа 2019 г., 22–25 августа 2023 г., г. Бердск); Всероссийской конференции молодых учёных по математическому моделированию и информационным технологиям (2019 г., г. Новосибирск); Всероссийской конференции «Преподавание информационных технологий в Российской Федерации» (2019 г., г. Новосибирск); 21-й Международной конференции «Геомодель 2019» (9–13 сентября 2019 г., г. Геленджик); Международной конференции «Марчуковские научные чтения – 2019» (2019 г., г. Новосибирск); Международной конференции «Applied Mathematics, Modeling and Simulation, AMMS 2021» (2021 г., онлайн); 24-м международном конгрессе «ISPRS Congress «Imaging today, foreseeing tomorrow» (2021 г., онлайн), 21-й международной научной конференции «GeoConference Surveying, Geology and Mining, Ecology and Management – SGEM 2022» (26 июня – 5 июля, 2021 г., онлайн), Устойчивое развитие особо охраняемых природных территорий (5–7 октября, 2023 г., г. Сочи), Цифровые технологии в горном деле (13–16 июня 2023 г., онлайн).

Диссертация подготовлена по результатам исследований, выполненных в рамках гранта, предоставленного в форме субсидии на выполнение Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг., а также на проведение крупных научных проектов по приоритетным направлениям научно-технологического развития в рамках подпрограммы «Фундаментальные научные исследования для долгосрочного развития и обеспечения конкурентоспособности общества и государства» государственной программы Российской Федерации «Научно-технологическое развитие Российской Федерации», проект «Социально-экономическое развитие Азиатской России на основе синергии транспортной доступности, системных знаний о природно-ресурсном потенциале, расширяющегося пространства межрегиональных взаимодействий», номер соглашения с Министерством науки

и высшего образования Российской Федерации № 075-15-2020-804 (внутренний номер гранта № 13.1902.21.0016), государственного задания Минобрнауки России по теме «Разработка теории и технологических решений контроля состояния защитных сооружений при перекачке нефтепродуктов методами активного дистанционного зондирования» (№ 0807-2020-0002), проекта № 121051900145-1 «Развитие научных основ, разработка и реализация новых безопасных и ресурсосберегающих физико-технической и физико-химической геотехнологий освоения месторождений твердых полезных ископаемых, склонных и опасных по газодинамическим явлениям и горным ударам», гранта Российского научного фонда (проект № 23-27-10057) и гранта НСО № р-60. «Разработка методики оценки состояния техногенно нарушенных земель Новосибирской области и прогнозирование направлений их альтернативного использования».

Результаты исследований внедрены в учебный процесс федерального государственного бюджетного образовательного учреждения высшего образования «Сибирский государственный университет геосистем и технологий» для дисциплин «Цифровая картография», «Современные научные и технологические направления в картографии и геоинформатике», «Базы пространственных данных» по направлению подготовки бакалавров 05.03.03 Картография и геоинформатика, «Теоретические концепции, проблемы и перспективы развития картографии и геоинформатики», «Системы интеллектуальной обработки пространственных данных» по направлению подготовки магистров 05.04.03 Картография и геоинформатика.

Публикации по теме диссертации. Основные теоретические положения и результаты исследований представлены в 37 научных публикациях, из них 14 – в изданиях, входящих в перечень российских рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора наук, 2 – свидетельства о регистрации программ для электронных вычислительных машин, 4 – патента РФ на изобретение, 12 статей – в изданиях, входящих в международные реферативные базы данных и системы цитирования Scopus и Web of Science.

Структура диссертации. Общий объем диссертации составляет 252 страницы машинописного текста. Диссертация состоит из введения, 4 разделов, заключения, списка сокращений и условных обозначений, списка литературы, включающего 253 наименования, содержит 6 таблиц, 60 рисунков и 5 приложений.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность темы исследований, показана степень разработанности проблемы, обоснованы и сформулированы цели и задачи, объект и предмет исследования, научная новизна и практическая значимость, приведены сведения об апробации предлагаемых решений, поставленных в диссертационной работе, ее структура и научные положения, выносимые на защиту.

В первом разделе «Анализ состояния проблемы использования технологий искусственного интеллекта в цифровой картографии» выполнено исследование современного состояния и практического опыта использования современных информационных технологий для автоматизации и роботизации процессов создания и обновления картографической и геоинформационной продукции, в том числе тех, которые используют технологии искусственного интеллекта с учетом специфики и проблем этих технологий при работе с пространственными данными. Анализ результатов исследований показал, что при большом объеме теоретических материалов, включая отечественные исследования по автоматизации процессов картографии и геоинформатики в целом, недостаточно проработаны вопросы использования технологий искусственного интеллекта на всех этапах картографического и геоинформационного производства.

Выделены основные технологии искусственного интеллекта, которые применяются при обработке пространственных данных: работа с естественными языками, извлечение, структурирование и использование знаний с помощью семантических структур, компьютерное зрение, самоорганизующиеся системы, компьютерное творчество, пространственно-временное прогнозирование.

Выявлены и сформулированы следующие основные проблемы использования систем искусственного интеллекта при решении задач картографии и геоинформатики:

- *необходимость комплексирования* решаемых в процессе обработки пространственных данных задач с использованием технологий ИИ;
- *сложность выбора* способа представления координат объектов геопространства как одного из параметров обрабатываемого элемента;
- *неприменимость* стандартных вариантов использования методов искусственного интеллекта для задач обработки и прогнозирования пространственных данных, что определяет проблему интерпретируемости и контроля результатов. Решением данной проблемы будет создание новых вариантов оценки или комбинирование существующих.

Во втором разделе «Формализация задач картографического и геоинформационного производства для их решения с применением технологий искусственного интеллекта» рассматривается схема типовых действий (рисунок 1), выполняемых при создании и обновлении картографических и геоинформационных продуктов и их автоматизация с помощью технологий искусственного интеллекта. В данном случае под формализацией понимается выделение отдельных элементарных процессов создания и обновления картографических и геоинформационных произведений, определение для них необходимых данных, требований к процессу обработки с учетом специфики конкретных методов и алгоритмов ИИ, выходных результатов и оценки качества.

С учетом особенностей перечисленных технологии ИИ и анализа современных процессов создания и обновления картографических произведений была составлена структурная схема типовых действий и соответствующих им технологий и направлений искусственного интеллекта (указаны в виде приведенных аббревиатур справа от блоков процесса, под ними указан средний потенциально достижимый процент автоматизируемости), выполняемых при создании и обновлении картографических и геоинформационных произведений, включающая в себя следующие блоки, приведенные на рисунке 1.

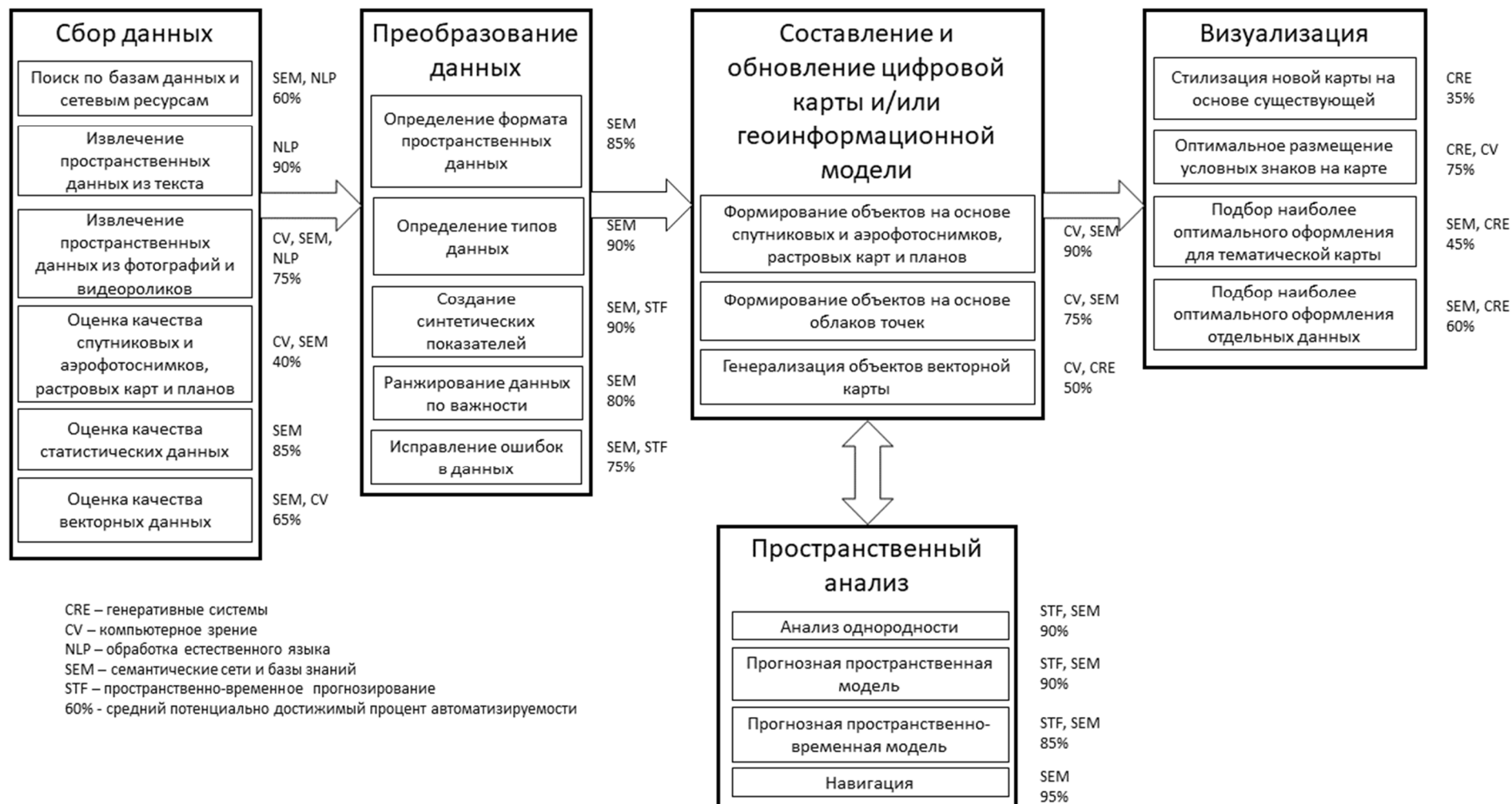


Рисунок 1 – Структурная схема связей процессов, выполняемых при создании и обновлении картографических и геоинформационных продуктов и их соотнесение с направлениями искусственного интеллекта

Для расчета степени автоматизации (P) предлагается использовать разработанную формулу:

$$P = (P_c + P_t + P_a + P_p + P_s)/5, \quad (1)$$

где P_c – процент покрытия задач этого типа; P_t – процент времени, затрачиваемого на ручную доработку; P_a – процент покрытия решения для текущего уровня развития геоинформационных технологий; P_p – процент точности обработки по сравнению с ручным вариантом; P_s – процент воспроизводимости. Числовые значения показателей получены на основе личных измерений и научных статей. На практике, процессы картографии и геоинформатики представляют различные комбинации отдельных атомарных задач, автоматизируемых с помощью технологий искусственного интеллекта.

Исходя из предложенной структурной схемы, предлагается отдельные элементы выделить в функциональные блоки, направленные на решение определенной задачи с использованием технологий ИИ, которые формируют концепцию «пространственного сервиса» (рисунок 2) для автоматизации подбора алгоритма / метода ИИ на основе графовой и больших языковых моделей, исходя из поставленной задачи.

Выбор блоков и порядок их выполнения формируется, исходя из текстового запроса на основе большой языковой модели, дообученной на нормативной документации и инструкциях, относящихся к цифровой картографии с учетом наиболее подходящих способов оценки итогового результата построения модели (пример матрицы критериев выбора способа оценки результата классификации по атрибутивным или пространственным данным, составленной в рамках исследования, приведен в таблице 1).

На основе выполнения блоков осуществляется обработка данных для решения поставленной задачи и формирование конвейеров обработки, реализуя принцип *posode*, а также формируются итоговые отчеты, содержащие, в том числе,

интерпретацию полученных моделей (базирующиеся на имеющихся способах и заданных при постановке задачи условиях, пример приведен в таблице 2).

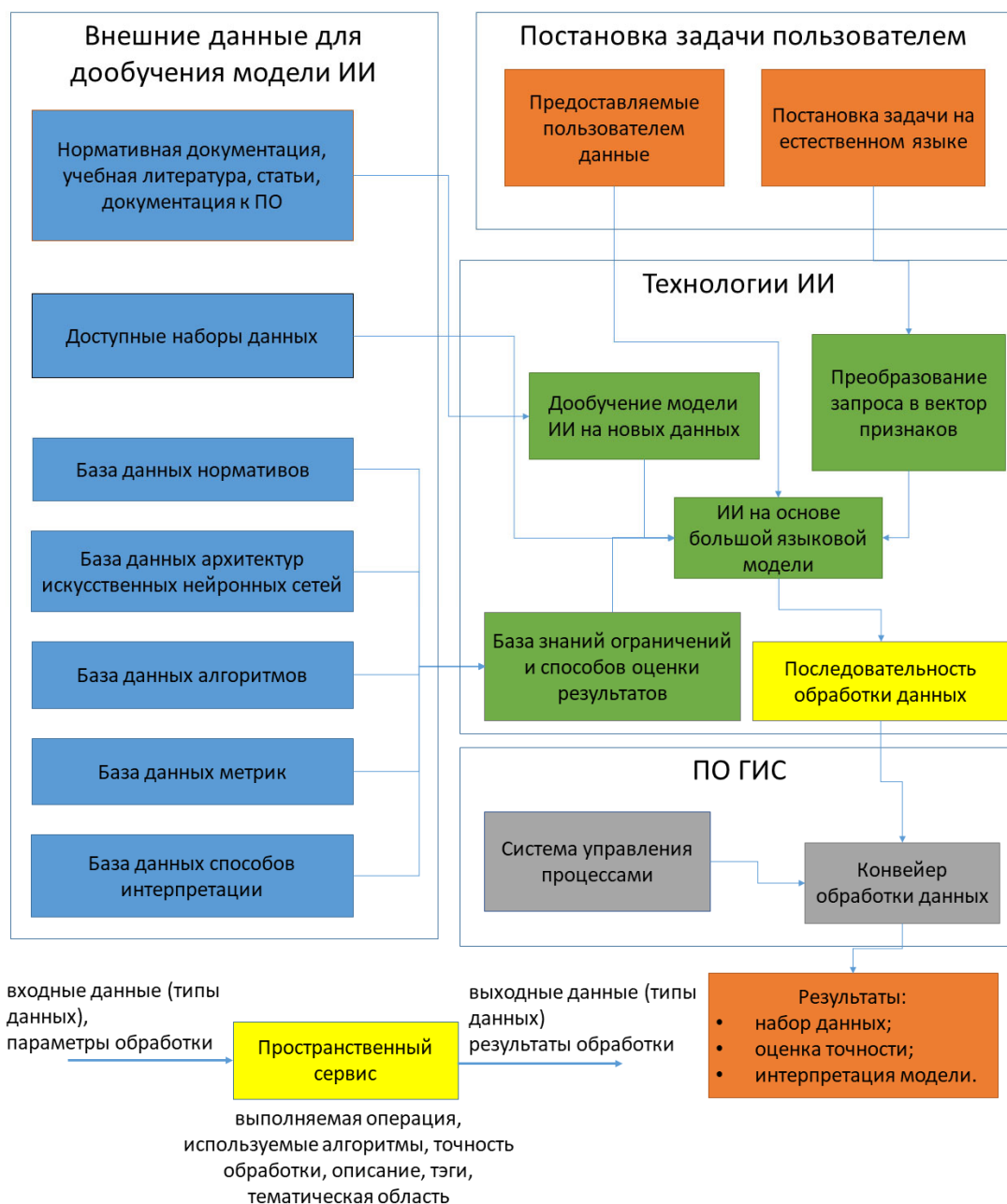


Рисунок 2 – Концепция «пространственного сервиса» для автоматизации подбора алгоритма / метода искусственного интеллекта, исходя из поставленной задачи

Таблица 1 – Фрагмент матрицы критериев выбора способа оценки результата классификации по атрибутивным или пространственным данным

Несбалансированность набора данных	Уровень ложноположительных результатов	Уровень ложноотрицательных результатов	Уровень положительных результатов среди всех истинных	Уровень положительных результатов среди всех объектов наруженных	Анализ пространственных объектов	Мультиклассовость	Способ оценки результата классификации по атрибутивным или пространственным данным
+	-	-	+	-	-	+	Confusion Matrix
-	+	-	-	-	-	+	False Positive Rate
-	-	+	-	-	-	+	False Negative Rate
-	-	-	+	-	-	-	Recall
-	-	-	-	+	-	-	Precision
+	-	-	-	-	+	+	P R curve
+	-	-	-	-	+	+	mIoU
+	-	-	-	-	+	+	Frequency weighted IoU

Таблица 2 – Фрагмент матрицы критериев выбора методов интерпретации моделей искусственного интеллекта

Комплексность объяснения работы модели	Скорость вычислений	Глобальность	Локальность	Возможность выявить гетерогенные отношения	Влияние корреляции признаков	Глобальные количественные оценки	Анализ графических данных	Метод интерпретации моделей искусственного интеллекта
Низкая	Высокая	Да	Нет	Нет	Да	Нет	Нет	Partial Dependence Plot
Низкая	Высокая	Да	Да	Да	Да	Нет	Нет	Individual Conditional Expectation
Низкая	Средняя	Да	Да	Да	Да	Да	Нет	Permuted Feature Importance
Средняя	Низкая	Да	Нет	Да	Нет	Да	Нет	Global Surrogate
Средняя	Низкая	Нет	Да	Да	Нет	Да	Нет	LIME
Высокая	Низкая	Да	Да	Да	Нет	Да	Нет	SHAP
Высокая	Низкая	Да	Да	Да	Нет	Да	Да	Attention Priority Map

Далее рассматриваются основные используемые элементы обработки и их особенности, начиная с входных данных. Способы представления (а также необ-

ходимость в дополнительной инженерии параметров пространственного положения) координат объектов геопространства, как одного из параметров обрабатываемого элемента алгоритмов искусственного интеллекта, требуется рассматривать, исходя из уровня детализации имеющихся данных, контекста решаемой задачи, предполагаемых требований к точности и скорости получения результата.

На основе анализа существующих решений и подходов автором предлагаются следующие варианты представления пространственных данных растровой и векторной моделей данных и облаков точек для использования в системах искусственного интеллекта. Основываясь на приведенном в работе описании вариантов представления пространственных данных, в рамках исследования составлена схема рекомендаций по выбору представления пространственных данных (рисунок 3, таблица 3), где голубым цветом выделены блоки с параметрами исходных данных, а фиолетовым – критерии выбора.

В соответствии со схемой на рисунке 1 видно, что в большей части процессов задействованы элементы извлечения, структурирования и использования знаний (и, более конкретно, геознаний, поскольку рассматриваются процессы обработки пространственных данных) с помощью семантических структур. Исходя из важности этого направления, автором предлагается представление геознаний (GK) как результата действий, получаемых на основе обработки и использования технологий искусственного интеллекта как элементов общей схемы (или структуры) данных (S) (включая необходимые преобразования), механизмов по формированию моделей (выводов) на их основе (P), способов оценки (M) и интерпретации полученного результата (I):

$$GK = S \cup P \cup M \cup I. \quad (2)$$

Схема данных (S) представляет собой описание типов данных (с подразделением на обязательные и возможные) (S_D) и необходимые преобразования (S_T).

$$S = S_D \cup S_T. \quad (3)$$

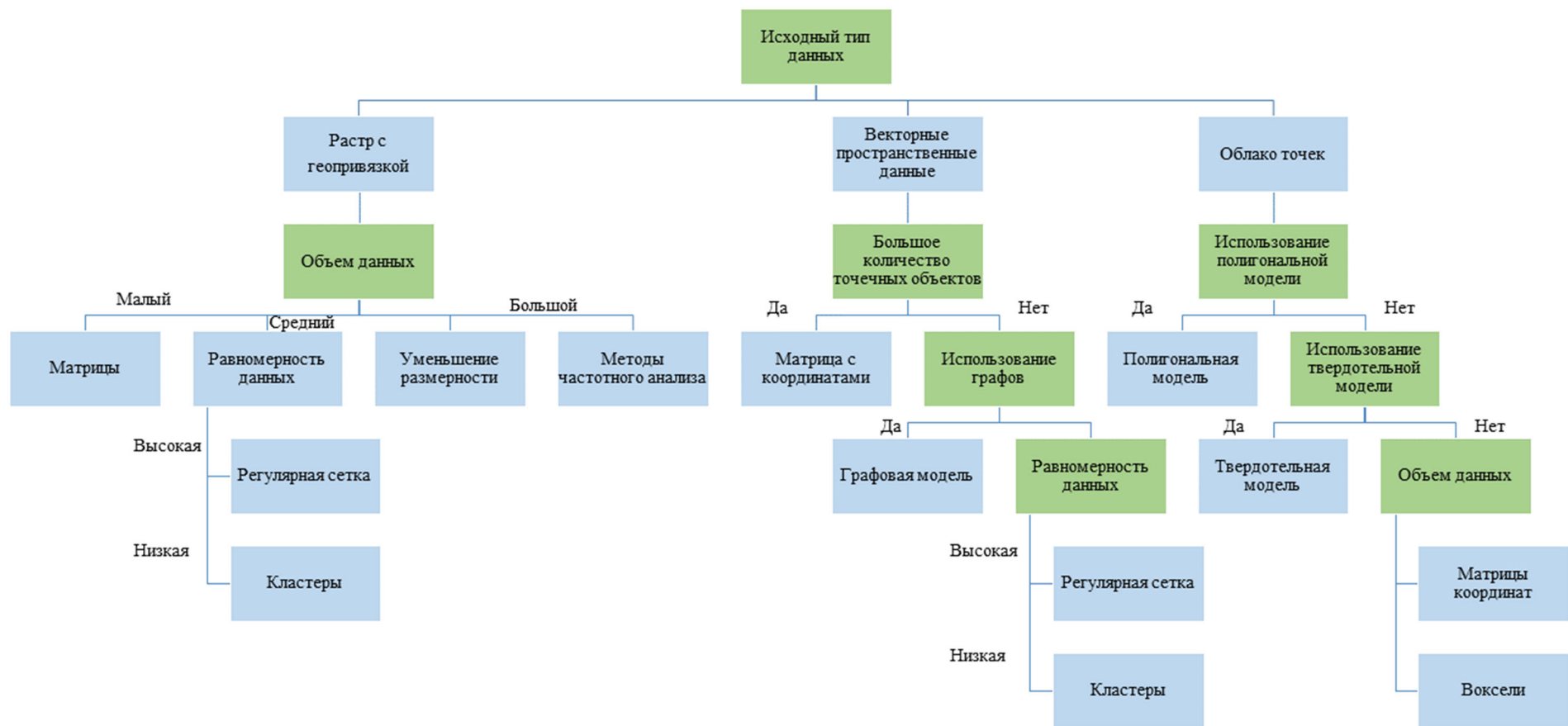


Рисунок 3 – Фрагмент схемы выбора представления пространственных данных для использования в системах, использующих технологии ИИ

Таблица 3 – Критерии оценки объема используемых пространственных данных

Оценка объема используемых пространственных данных	Соотношение площади блока данных с площадью анализируемой территории	Количество записей	Степень структурированности данных	Необходимость распределенной обработки	Средства хранения данных
Малый	Менее 1/100 000	Тысячи	Максимальная	Нет	Файлы
Средний	От 1/100 000 до 1/10 000 000	Миллионы	Высокая	Нет	СУБД, DWH
Большой	Более 1/10 000 000	Миллиарды и выше	Произвольная	Да	Системы распределенного хранения, LakeHouse

Исходные данные (D) представляют собой совокупность пространственно координированных данных (G) и статистических данных (S), определенных на один или более момент времени ($1 \dots n$). Необходимые преобразования (S_T) составляют очень важную и не менее объемную часть схемы данных (S) для того, чтобы сформировать единое геопространство из всего имеющегося массива данных

$$D = DG_{T1} \cup DS_{T1}, DG_{T2} \cup DS_{T2}, \dots, DG_{Tn} \cup DS_{Tn}. \quad (4)$$

Компонент обработки может представлять собой как простые условия, так и сложные математические модели (включая нейронные сети). При этом они могут быть ансамблированы, как и в случае обработки данных любых других видов. Из всех элементов геознания этот компонент имеет наибольшую вариативность, поскольку может быть реализован десятками алгоритмов, методов и технологий. Поскольку в геознаниях нет непосредственно данных, а они поступают извне, то для каждого нового набора данных нужны способы оценки качества и достоверности полученного результата (M). Кроме непосредственной оценки результатов обработки, важно интерпретировать (I) в варианте, понятном для человека, как именно происходит процесс формирования выводов (P). В предлагаемом варианте геознания не обязательно должны иметь графическое представление в виде карты, которая получается на основе результатов обработки с использованием

геознаний, представляющих собой также пространственные и статистические данные, но с большей степенью обобщенности.

$$GK: D \rightarrow D_{GK}. \quad (5)$$

Описанные компоненты геознаний далее используются при описании процессов использования технологий искусственного интеллекта для решения задач цифровой картографии и геоинформатики.

В третьем разделе «Методология использования технологий искусственного интеллекта при решении задач картографии и геоинформатики» исследуются задачи цифровой картографии и геоинформатики на основе процессов создания и использования картографических и геоинформационных продуктов, рассмотренных во втором разделе. Для каждого процесса приводится анализ входящих в него элементов и предлагаются методологические решения их выполнения с помощью технологий искусственного интеллекта с учетом оптимальности достижения итогового результата. В работе предлагается следующее методологическое решение вычисления критерия оптимальности (O_{sd} , меньшее значение – лучше, пример в таблице 4):

$$O_{sd} = \sum_1^n O_n * \sum_1^n OI_n * \sum_1^n H_n * V_{td} * \frac{1}{M} * \sum_1^n T_n * \sum_1^n TI_n, \quad (6)$$

где O_n – сложность обучения для i -й математической модели ($1 \dots n$); OI_n – сложность обработки (инференса) модели; H_n – пространственная сложность (к памяти аппаратной части); V_{td} – объем тренировочных данных; M – точность по выбранной метрике; T – время, затрачиваемое на преобразования пространственных данных; TI – время, затрачиваемое на интерпретацию модели.

Таблица 4 – Сводные параметры требований к ресурсам алгоритмов машинного обучения (f – число объектов; m – число атрибутов; k – число соседей; t – число деревьев; l – число слоев; n – число нейронов; e – число итераций)

Алгоритм	Критерий оптимальности	Сложность обучения	Сложность обработки (инференса)	Пространственная сложность
Линейная регрессия	$5,12 e^{15}$	$f*m^2+m^3$	m	m
Логистическая регрессия	$5,44 e^{12}$	$f*m$	m	m
k -ближайших соседей	$2,69 e^{12}$	$f*m*k$	$n*m$	$n*m$
Случайный лес	$2,10 e^{12}$	$t*f*\log(f)*m$	$m*t$	$t*\text{глубина дерева}$
Нейронная сеть	$2,28 e^{12}$	l^2*n^2*e	$l*n$	l^2*n^2

Извлечение пространственных данных из текста

В работе автором предлагается методологическое решение в виде технологической схемы построения и оценки моделей извлечения именованных сущностей (NER) с использованием технологий ИИ, ориентированное на обработку текстов, содержащих наименования географических объектов, а также дополнительные категории именованных сущностей, позволяющие с большей степенью автоматизировать (точность порядка 87 %) классифицирование наименований объектов (рисунок 4).

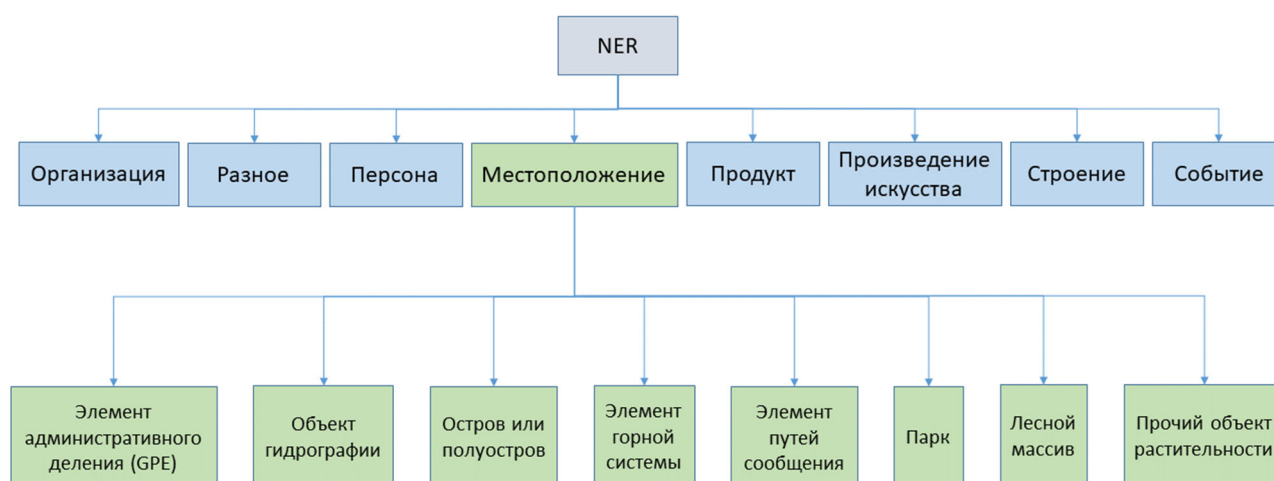


Рисунок 4 – Схема элементов классификации NER с расширенным обозначением географических объектов

Поиск по базам данных и сетевым ресурсам

Переходя на более высокий уровень поиска и извлечения данных в работе предлагаются методологические решения по семантическому пространственному поиску и дается следующее его определение. Семантический пространственный поиск – это поиск, который в дополнение ключевым словам использует пространственные базы данных и знаний объектов реального мира, методы нейронных сетей для обработки запросов и выдачи результата (с учетом смысла запросов и контента) с адаптацией их на основе новых данных и поисковых запросов пользователей. То есть для этого используются знания реального мира, чтобы определить намерения пользователя геоинформационной и картографической продукции.

Извлечение пространственных данных из фотографий и видеороликов

Поскольку в результате использования семантического поиска, могут быть получены не только структурированные, но и полуструктурированные данные, то, исходя из этого, рассматривается задача и предлагается методологическое решение для извлечения данных о геолокации из фотографий и видео, состоящее в том, чтобы определять местоположение (координаты) изображения или видео с максимально возможной точностью (до 100 квадратных метров), используя только значения и взаимоположение пикселей, без метаданных камеры (с обеспечением 34, 50, 69, 83 % точности для города, региона, страны и континента соответственно). В работе предлагается следующая схема выбора метода на основе технологий искусственного интеллекта для определения местоположения фотографии. В цифровом варианте данная схема представляет собой семантическую графовую структуру и предназначена для применения в экспертных системах и других программных приложениях, использующих технологии искусственного интеллекта (рисунок 5).

Поэтому требуется уточнить, что эту структуру невозможно полностью представить визуально (ввиду ее многомерности), и на тех рисунках, где приводятся схемы выбора способов представления данных, методов, алгоритмов, она

представлена фрагментарно. Полный вариант размещен по адресу https://github.com/AlexeyKW/Geo_AI.

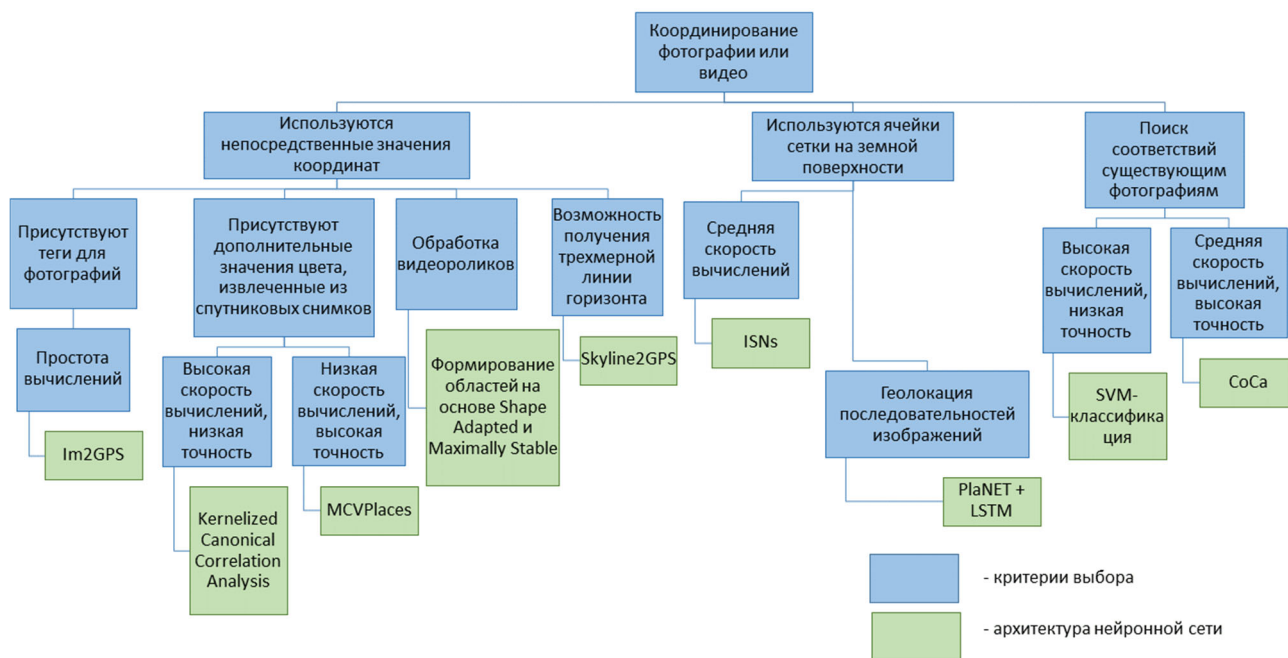


Рисунок 5 – Фрагмент схемы выбора метода определения местоположения фотографии

Дополнительным способом уточнения положения фотографии может быть извлечение и распознавание текстов, попавших в объектив камеры. Для увеличения качества автоматизированного выявления областей, возможно, содержащих буквы и цифры, автором предлагается методологическое решение в виде последовательности действий на основе использования ансамблирования обученного каскада Хаара и гистограммного анализа областей и систем оптического распознавания текстов.

Формирование объектов на основе спутниковых и аэрофотоснимков, растровых карт и планов

Как развитие методов обработки растровых изображений, далее рассматривается использование технологий искусственного интеллекта для их семантической сегментации. На основе анализа существующих архитектур искусственных нейронных сетей автором предлагается методологическое решение для выбора

метода и особенностей архитектуры нейронной сети для семантической сегментации растровых геоизображений. Также, на основе анализа существующих подходов, базирующихся на технологиях искусственного интеллекта, автором предлагается несколько классификаций архитектур нейронных сетей, использующихся для обработки трехмерных радиолокационных данных, облаков точек, мониторинга на основе трехмерных данных.

Определение формата и типов пространственных данных

С точки зрения автоматизации определения типов данных атрибутов объектов цифровой карты предложено методологическое решение, выявляющее типы данных (на основе IEEE Std 1320.2-1998 и DBPedia Semantic types), полагаясь на семантическое сходство между таблицами и их столбцами. Технологическая схема подготовки и использования математической модели для определения типов данных (с учетом наличия пространственных данных) с помощью технологий искусственного интеллекта (с обеспечением точности в 83 %) представлена на рисунке 6.

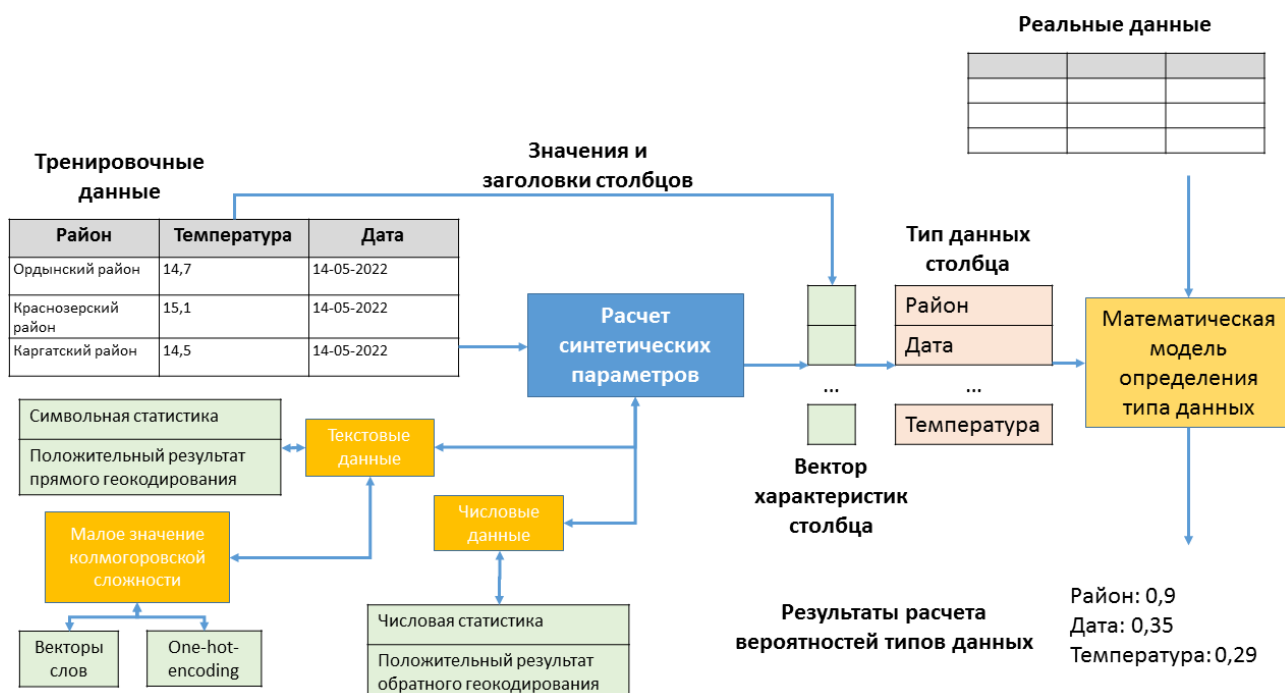


Рисунок 6 – Схема подготовки и использования математической модели для определения типов данных с помощью технологий ИИ

Стилизация карты

Поскольку неотъемлемой частью картографии является визуализация пространственных и атрибутивных данных, то в работе предлагается методологическое решение на основе использования архитектур генеративно-состязательных сетей (GAN) для автоматической стилизации картографических изображений (рисунок 7).

Базовое изображение	Результат обработки	Образец исходного стиля
		
		

Рисунок 7 – Примеры работы генеративно-состязательной сети

Анализ однородности

С точки зрения визуализации также рассматривается такая особенность, что современные данные часто являются многомерными в варианте большого числа семантических характеристик, особенно если речь идет о пространственных данных. В этом случае предлагается методологическое решение, суть которого заключается в использовании алгоритмов и методов снижения размерности для отбора, либо синтетического создания, необходимого для визуализации на карте количества характеристик. В заключении третьего раздела рассматриваются эле-

менты пространственного анализа, но поскольку, они, в большей степени, относятся с области геоинформатики, в части оценки имеющихся пространственных данных, то основное внимание уделяется оценке имеющихся пространственных данных, также кратко даны рекомендации по выбору метода пространственно-временного прогнозирования на основе технологий искусственного интеллекта. Перечисленные в третьем разделе методологические решения представляют собой новую методологию использования технологий искусственного интеллекта на отдельных этапах картографического производства.

В четвертом разделе объясняется необходимость и предлагаются решения по качественной оценке работы программных продуктов, основанных на технологиях искусственного интеллекта, и приводятся примеры апробации разработанных концептуальных, методических, технических и технологических решений при построении оптимальных конвейеров обработки пространственных данных в процессе решения определенной задачи цифровой картографии.

Всесторонняя и качественная оценка результатов работы программных решений очень важна, поскольку при использовании технологий искусственного интеллекта происходит зауживание взглядов к решению проблем, ориентирование на использование только этих технологий ввиду того, что они дают гарантированный результат и являются универсальным решением, формируя убеждения о ненужности и малой значимости экспертного мнения и традиционных решений. Для уменьшения негативного влияния этой проблемы нужно максимально критически подходить к результатам работы программных решений, построенных на основе технологий искусственного интеллекта, использовать всесторонние критерии анализа качества результатов и обязательно сравнивать их с традиционными подходами. В первом подразделе четвертого раздела приводятся фрагменты схем критериев выбора способа оценки для задач классификации, семантической сегментации (рисунок 8), прогнозирования количественных значений, качества работы семантического пространственного поиска, работы генеративной модели при стилизации картографических произведений.

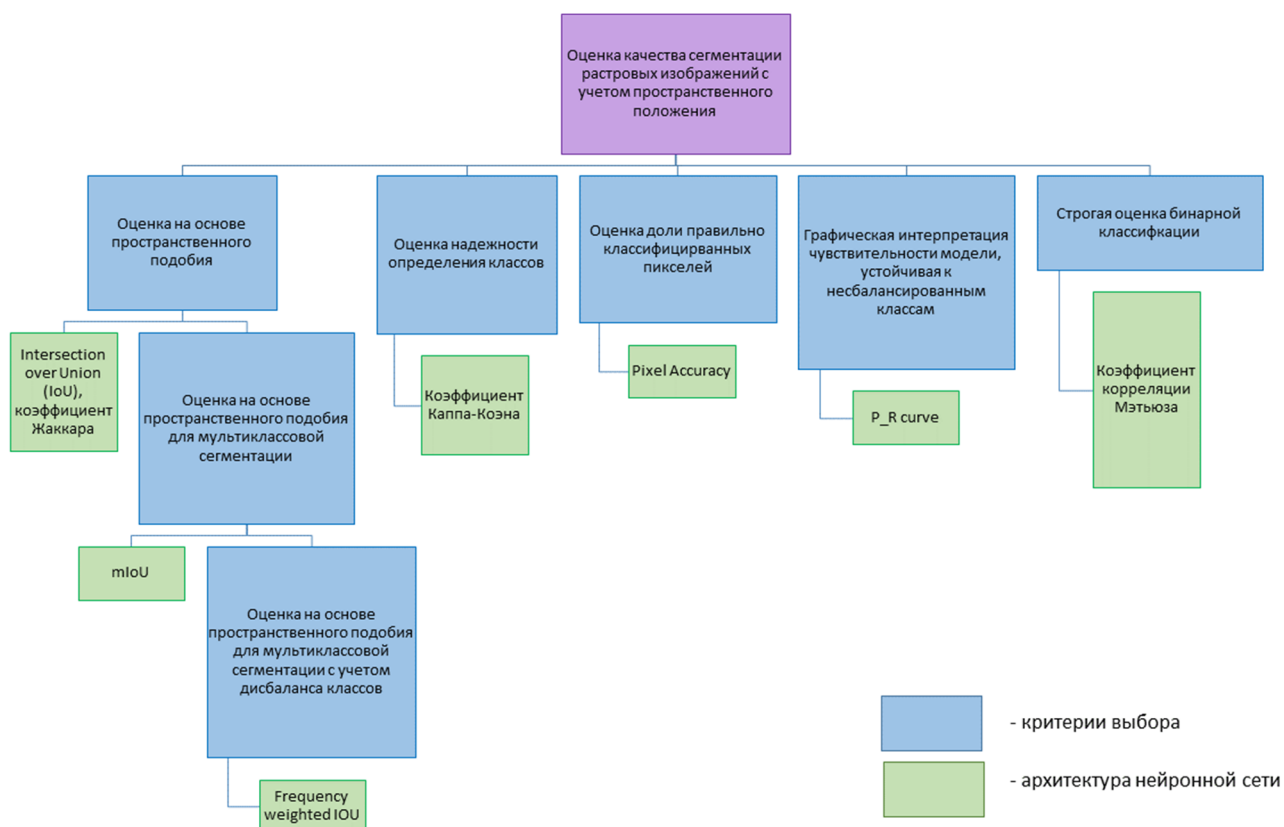


Рисунок 8 – Фрагмент схемы выбора способа оценки качества семантической сегментации растровых изображений

На основе базовых способов оценки качества классификации автором предлагается следующее методологическое решение по использованию количественных «строгих» (7)–(10) и «мягких» (11)–(14) оценок при поиске наименований пространственных объектов в тексте

$$Accuracy_{strict} = \frac{c3 + cw}{c3 + cw + c0 + c1 + c2 + c'} \quad (7)$$

$$Precision_{strict} = \frac{c3}{c3 + c2 + c'} \quad (8)$$

$$Recall_{strict} = \frac{c3}{c3 + c0 + c1} \quad (9)$$

$$F1_{strict} = 2 * \frac{Precision_{strict} * Recall_{strict}}{Precision_{strict} + Recall_{strict}} \quad (10)$$

$$Accuracy_{soft} = \frac{c2 + c3 + cw}{c3 + cw + c0 + c1 + c2 + c'} \quad (11)$$

$$Precision_{soft} = \frac{c2 + c3}{c3 + c2 + c'} \quad (12)$$

$$Recall_{soft} = \frac{c2 + c3}{c3 + c0 + c1} \quad (13)$$

$$F1_{soft} = 2 * \frac{Precision_{soft} * Recall_{soft}}{Precision_{soft} + Recall_{soft}} \quad (14)$$

где c – количество ошибочно найденных геообъектов (ложноположительный вариант); $c0$ – количество пропущенных геообъектов; $c1$ – количество геообъектов, отнесенных к другому классу; $c2$ – количество частично найденных геообъектов; $c3$ – количество полностью корректно найденных геообъектов; cw – общее число слов в тексте за исключением геообъектов.

Для оценки точности получения координат фотографии, при определении местоположения устройства съемки, автором предлагается ввести дополнительный количественный способ оценки на основе плотности населения и плотности фотографий в этом месте. Для расчета коэффициента плотности населения и фотографий используется нормированная плотность населения в месте потенциального определения местоположения и количество фотографий в радиусе, двукратно превышающем расстояние до ближайшей фотографии относительно самого анализируемого изображения) по формуле

$$pd_i = \frac{PD_i}{PD_{max}} * PC_i, \quad (15)$$

где PD_i – нормированная плотность населения в месте расположения i -го изображения; PD_i – плотность населения в месте съемки изображений i ; PD_{max} – максимальная плотность населения Земли; PC_i – количество фотографий в радиусе 500 м (включая само анализируемое изображение).

Предлагаемый в работе расчет погрешности определения координат производится по следующей формуле:

$$error = \sqrt{(1 + 0,1 \cdot pd_i) \cdot d^2}, \quad (16)$$

где d – расстояние между предсказанными и истинными географическими координатами изображения.

Важность интерпретируемости и объяснимости складывается из следующих факторов: степень доверия пользователя к результатам работы математической модели, возможность исследований относительно того, как улучшить работу модели, обеспечение безопасности и возможность формирования состязательных примеров (adversarial examples), получение новых знаний на основе построенных моделей, социальная, философская и этическая составляющие. Для этого направления рассматриваются существующие подходы (LIME, SHAP, Feature Importance, карты градиентов и т. д.) и делаются рекомендации для модельезависимых и модельезависимых методов.

Предложенные методологические решения, касающиеся распределенной обработки и построения прогностических математических моделей, применялись при сравнении скорости обработки различных данных дистанционного зондирования для одной и той же территории с использованием традиционного подхода и технологий распределенной обработки на основе комбинации функций Apache Airflow и Apache NiFi. Описанный эксперимент показал увеличение скорости обработки данных в несколько раз, в зависимости от числа узлов обработки данных.

В порядке апробации предложенных в работе методологических решений нами выполнен ряд проектов по использованию методов и технологий искусственного интеллекта для автоматизации решения задач цифровой картографии.

Как пример применения технологий искусственного интеллекта для построения пространственных имитационных моделей была разработана топографо-климатическая компонента имитационной модели каскада гидроэлектростанций. Этот объект интересен тем, что учитывает не только различные алгоритмы управления этим объектом, но и изменение входных параметров, обусловленное природными условиями, которые, в свою очередь, можно анализировать и стро-

ить прогнозы поведения на основе пространственных данных (спутниковая и БАС съемка, наземная геодезическая съемка, значения с метеостанций и т. д.).

Методы искусственного интеллекта для извлечения пространственной информации из разнородных данных были опробованы на примере использования данных, полученных из социальных сетей, в качестве источника информации об угрозе возникновения чрезвычайных ситуаций. Для апробации было выполнено построение карт по принципам оперативного картографирования с целью выявления повреждений и определения состояния и доступности дорожной сети на основе данных спутниковой съемки, БАС и радиолокационной съемки с помощью различных методов анализа изображений. В качестве рассматриваемой территории была выбрана Индонезия в момент цунами 28 сентября 2018 г. Используемые наборы данных относятся к открытым данным следующих источников: Sentinel-1, Sentinel-2, ALOS-2, Landsat-8. Были собраны изображения, относящиеся к поврежденным, но еще доступным для внедорожного транспорта дорогам, а также изображения обычных дорог. Результаты обработки показали, что разработанный подход, основанный на полностью сверточных нейронных сетях с использованием нескольких источников данных оптического дистанционного зондирования и данных SAR, подходит для нужд оценки и картирования дорог при управлении чрезвычайными ситуациями, в случае отсутствия данных картографирования дорог во время чрезвычайной ситуации. Он показывает очень хорошие результаты, если используются данные БАС, предоставляя граф дорог с топологической достоверностью 75 и 68 % и точностью классификации дорог 93 и 73 % соответственно для неповрежденных дорог и частично доступных дорог.

Использование методов искусственного интеллекта в области построения прогнозных математических моделей и тематических карт на их основе для экологических систем и явлений апробировалось на следующих наборах данных: численность moskitov *Aedes aegypti* на территории западной и прибрежной частей Кении, количество заражений лихорадкой Денге на территории Пуэрто-Рико и Перу, физическое состояние деревьев в низинной тропической роще на

территории Пуэрто-Рико, уровень воды в бассейне реки Соан, Пакистан, изменения контуров техногенно нарушенных территорий Новосибирской области по данным мультиспектральных разновременных спутниковых снимков.

В качестве примера работы разработанной методологии приведены результаты работы конвейера выделения контуров техногенно нарушенных территорий по данным спутниковых изображений. В качестве апробации технических решений созданы конвейеры обработки данных на основе систем управления процессами, в том числе используемые для автоматизации наполнения и обновления баз данных «Техногенно-измененные земли при добыче и использовании твердых полезных ископаемых на территории Новосибирской области», «База знаний показателей объектов недвижимости на территорию Новосибирской области по данным из открытых источников», построения прогнозных карт распространения инфекционных заболеваний, параметров объектов гидрографии по климатическим параметрам, формирования пространственных баз данных на основе неструктурированных текстов, анализа качества исходных данных.

ЗАКЛЮЧЕНИЕ

В результате проведенного диссертационного исследования реализовано новое комплексное направление в картографии, связанное с применением технологий искусственного интеллекта для автоматизации решения задач цифровой картографии и геоинформатики.

Поставленная цель – разработка теории и методологии использования технологий искусственного интеллекта для решений задач цифровой картографии – достигнута, что позволит обеспечивать безопасность и устойчивое развитие экономики Российской Федерации. Основные научные и практические результаты, полученные в ходе исследования, заключаются в следующем.

1 На основе анализа объектов исследования разработана теория и сформирована общая концепция использования технологий искусственного интеллекта

для решения задач цифровой картографии, позволяющая обеспечить от 15 до 100 % доли автоматизации как комплексных процессов цифровой картографии, так и отдельных задач обработки пространственных и статистических данных.

2 Исследована степень изученности, алгоритмы, методы и технические решения, способы применения и ограничения использования технологий искусственного интеллекта, позволяющие до 50 % сократить время на выбор оптимального технического решения применительно к определенной задаче цифровой картографии.

3 Разработаны принципы и понятийный аппарат, обеспечивающие научно обоснованное использование технологий искусственного интеллекта для автоматизации и роботизации процессов картографирования.

4 Разработана методология использования технологий искусственного интеллекта, базирующаяся на теоретически обоснованном логическом аппарате выбора и разработки: способа представления пространственных данных, их обработки определенными алгоритмами и методами, критериев оценки качества результата базирующихся на технологиях искусственного интеллекта, позволяющая выполнять автоматизированное (вплоть до 80 %) построение и реализацию оптимального конвейера обработки пространственных данных при решении определенной задачи цифровой картографии.

5 Реализованы технические решения в виде модулей для геоинформационных систем, позволяющие использовать технологические элементы работы для автоматизации процессов создания и редактирования картографических произведений, а также коллективных многоплановых исследований технологий искусственного интеллекта для задач обработки пространственных данных (обеспечивающих от 15 до 100 % автоматизации процессов обработки пространственных и статистических данных по текстовым запросам на естественном языке, до 83 % автоматизировать процессы картографирования по текстовым данным, автоматизировать подбор источников данных для построения карт

и планов до 35 %, определять семантические типы данных атрибутов векторных картографических данных с точностью до 79 %).

6 Разработан комплекс баз данных и знаний для экспертной системы подбора оптимальной (обеспечивающий уменьшение времени выполнения процессов вплоть до 45 %) последовательности технологических этапов и обработки пространственных данных системами искусственного интеллекта при решении задач цифровой картографии.

Таким образом, поставленная цель и сформулированные задачи решены. В результате исследований получены новые научно обоснованные технологические и технические решения по автоматизации процессов цифровой картографии, обеспечивающие от 15 до 100 % автоматизации отдельных задач и процессов обработки пространственных и статистических данных по текстовым запросам на естественном языке. Предложены решения, позволяющие на 83 % автоматизировать (с точностью порядка 87 %) процессы картографирования по текстовым данным, автоматизировать подбор источников данных для построения карт и планов на 35 %, определять местоположение изображения или видео без использования метаданных камеры (с обеспечением 34, 50, 69, 83 % точности для уровней города, региона, страны и континента соответственно), определять семантические типы данных атрибутов векторных картографических данных с точностью в 79 %, оптимизировать объемы хранимых растровых пространственных данных от 20 до 65 % за счет алгоритмов снижения размерности.

Результаты исследований могут быть использованы при картографическом и геоинформационном производстве, в научных исследованиях, связанных с обработкой и анализом пространственных данных с точки зрения автоматизации процессов.

Перспективы дальнейших исследований направлены на расширение функциональных возможностей программного обеспечения, увеличение степени автоматизации и упрощения интеграции технологий искусственного интеллекта

в научные исследования и производство картографической и геоинформационной продукции.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

1 Колесников, А. А. Возможности использования web-технологий для визуализации данных, получаемых с помощью активных методов дистанционного зондирования / А. А. Колесников, Д. В. Грищенко. – Текст : непосредственный // Вестник СГУГиТ. – 2020. – Т. 25, № 4. – С. 130–137.

2 Предпроектные исследования гео моделирования контролируемых показателей и составления кадастровой карты государственного реестра объектов негативного воздействия на окружающую среду / О. В. Алайская, Ш. Чжэньфэн, Х. Энамул, А. А. Колесников, А. М. Портнов. – Текст : непосредственный // Естественные и технические науки. – 2020. – № 6 (144). – С. 110–120.

3 Применение ГИС-технологий для повышения эффективности имитационного моделирования каскадов ГЭС / А. Ю. Арестова, С. В. Митрофанов, А. Г. Русина, А. А. Колесников. – Текст : непосредственный // Журнал Сибирского федерального университета. Серия: Техника и технологии. – 2020. – Т. 13, № 6. – С. 732–744.

4 Молокина, Т. С. Анализ состояния и перспективы развития визуализации пространственных данных / Т. С. Молокина, А. А. Колесников. – Текст : непосредственный // Вестник СГУГиТ. – 2021. – Т. 26, № 4. – С. 73–82.

5 Колесников, А. А. Анализ методов и средств искусственного интеллекта для анализа и интерпретации данных активного дистанционного зондирования / А. А. Колесников. – Текст : непосредственный // Вестник СГУГиТ. – 2022. – Т. 27, № 3. – С. 74–94. – DOI 10.33764/2411-1759-2022-27-3-74-94.

6 Колесников, А. А. Возможности NOSQL СУБД для обработки пространственных данных / А. А. Колесников. – Текст : непосредственный // Вестник

СГУГиТ. – 2022. – Т. 27, № 3. – С. 95–106. – DOI 10.33764/2411-1759-2022-27-3-95-106.

7 Использование геопространственных данных для оценки состояния техногенно нарушенных земель / Н. С. Косарев, А. А. Колесников, А. В. Резник [и др.]. – Текст : непосредственный // Физико-технические проблемы разработки полезных ископаемых. – 2023. – № 6. – С. 190–197. – DOI 10.15372/FTPRPI20230617.

8 Колесников, А. А. Современные подходы к автоматизации операций по редактированию пространственных данных / А. А. Колесников. – Текст : непосредственный // Геодезия и картография. – 2023. – Т. 84, № 4. – С. 39–49. – DOI 10.22389/0016-7126-2023-994-4-39-49.

9 Колесников, А. А. Технические решения для создания картографических баз данных на основе неструктурированного текста с использованием микросервисной архитектуры и систем обработки естественного языка / А. А. Колесников. – Текст : непосредственный // Информация и космос. – 2023. – № 2. – С. 132–136.

10 О необходимости комплексной геоэкологической оценки техногенно нарушенных горными работами земель / В. Л. Гаврилов, Н. А. Немова, А. В. Резник [и др.]. – Текст : непосредственный // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2023. – Т. 334, № 10. – С. 76–87. – DOI 10.18799/24131830/2023/10/4212.

11 Создание базы данных техногенно-нарушенных территорий Новосибирской области / А. А. Колесников, Н. С. Косарев, Н. А. Немова [и др.]. – Текст : непосредственный // Вестник СГУГиТ (Сибирского государственного университета геосистем и технологий). – 2023. – Т. 28, № 5. – С. 80–92. – DOI 10.33764/2411-1759-2023-28-5-80-92.

12 Автоматизация определения контуров техногенно нарушенных территорий по данным открытой спутниковой съемки / А. А. Колесников, Н. С. Косарев, А. В. Резник, Н. А. Немова, А. М. Астапов, М. К. Кропачева. – Текст : непосред-

ственный // Геодезия и картография. – 2024. – № 11. – С. 25–34. – DOI 10.22389/0016-7126-2024-1013-11-25-34.

13 Колесников, А. А. Использование больших языковых моделей в геоинформационных технологиях / А. А. Колесников. – Текст : непосредственный // Известия вузов. Геодезия и аэрофотосъемка. – 2024. – Т. 68, № 1. – С. 33–43. – DOI 10.30533/GiA-2024-003.

14 Колесников, А. А. Сравнительный анализ влияния компонентов мультиспектрального снимка на точность автоматизированного дешифрирования объектов гидрографии / А. А. Колесников. – Текст : непосредственный // Информация и космос. – 2024. – № 2. – С. 141–144.

15 Патент на изобретение. Способ определения местоположения, координат точек, геометрических и семантических характеристик картографических объектов в интерактивном режиме при работе с традиционной картой в условиях отсутствия связи для передачи цифровых данных : 2706465 С2, 19.11.2019 : заявл. № 2019111792 от 18.04.2019Д. / Лисицкий Д. В., Комиссарова Е. В., Колесников А. А. – Текст : непосредственный.

16 Патент на изобретение. Способ создания автономного источника геопространственной информации и использования его в интерактивном режиме в условиях отсутствия связи для передачи цифровых данных : 2752606 С2, 29.07.2021 : заявл. № 2019120894 от 02.07.2019 / Лисицкий Д. В., Комиссарова Е. В., Колесников А. А. – Текст : непосредственный.

17 Патент на изобретение. Способ определения местоположения на традиционной карте координат точек, геометрических и семантических характеристик картографических объектов в интерактивном режиме при работе с традиционной картой в условиях отсутствия связи для передачи цифровых данных : 2757066 С2, 11.10.2021 : заявл. № 2019116623 от 29.05.2019 / Лисицкий Д. В., Комиссарова Е. В., Колесников А. А. – Текст : непосредственный.

18 Патент на изобретение. Способ получения, обработки, отображения и интерпретации геопространственных данных для кластеризации неоднородности

техногенно измененных территорий : 2806406 С1 Российская Федерация, МПК G01С 11/04, G06Т 17/05: заявл. № 2022125547 от 30.09.2022 / А. В. Резник, В. Л. Гаврилов, Н. А. Немова [и др.] ; заявитель Федеральное государственное бюджетное учреждение науки Институт горного дела им. Н. А. Чинакала Сибирского отделения Российской академии наук. – Текст : непосредственный.

19 Свидетельство о регистрации программы для ЭВМ : QSPARQL : RU 2018611179, 24.01.2018 : заявл. № 2017662399 от 30.11.2017 / Колесников А. А., Кикин П. М. – Текст : непосредственный.

20 Свидетельство о государственной регистрации программы для ЭВМ № 2023686249 Российская Федерация. Программа для извлечения и координирования географических объектов из неструктурированного текста : № 2023686275 : заявл. 05.12.2023 : опубл. 05.12.2023 / Колесников А. А. ; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Сибирский государственный университет геосистем и технологий». – Текст : непосредственный.

21 Kolesnikov, A. A. Forecasting the distribution of diseases in tropical zones using machine learning methods / A. A. Kolesnikov, P. M. Kikin. – Текст : непосредственный // CEUR Workshop Proceedings. SDM 2019 – Proceedings of the All-Russian Conference «Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes», 2019. – P. 371–376.

22 Kikin, P. M. Use of machine learning techniques for rapid detection, assessment and mapping the impact of disasters on transport infrastructure / P. M. Kikin, A. A. Kolesnikov, A. M. Portnov. – Текст : непосредственный // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives, 2019. – P. 195–200.

23 Kolesnikov, A. A. Diseases spread prediction in tropical areas by machine learning methods ensembling and spatial analysis techniques / A. A. Kolesnikov, P. M. Kikin, A. M. Portnov. – Текст : непосредственный // International Archives of

the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives, 2019. – P. 221–226.

24 Kikin, P. M. Social media data processing and analysis by means of machine learning for rapid detection, assessment and mapping the impact of disasters / P. M. Kikin, A. A. Kolesnikov, E. A. Panidi. – Текст : непосредственный // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives. – 2020. – 24. – P. 1237–1241.

25 Cloud-desktop remote sensing data management to ensure time series analysis, integration of QGIS and Google Earth Engine / E. Panidi, I. Rykin, P. Kikin, A. Kolesnikov. – Текст : непосредственный // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives. – 2020. – 24. – P. 553–557.

26 Системы обработки естественного языка для извлечения данных и картографирования на основе неструктурированных блоков текста / А. А. Колесников, П. М. Кикин, Дж. Нико, Е. В. Комиссарова. – Текст : непосредственный // ИнтерКарто. ИнтерГИС. – 2020. – Т. 26, № 1. – С. 375–384.

27 Анализ и прогнозирование пространственно-временных экологических показателей с использованием методов машинного обучения / П. М. Кикин, А. А. Колесников, А. М. Портнов, Д. В. Грищенко. – Текст : непосредственный // ИнтерКарто. ИнтерГИС. – 2020. – Т. 26, № 3. – С. 53–61.

28 Integrating GIS technologies in hydro power plant cascade simulation model A. Khalyasmaa, S. Eroshenko, A. Arestova, S. Mitrofanov, A. Rusina, A. Kolesnikov. – Текст : непосредственный // E3S Web of Conferences. Серия: 2020 The 3rd International Conference on Renewable Energy and Environment Engineering, REEE 2020. – 2020. – 3: 02006.

29 Kolesnikov, A. Building a model for predicting the water level in a river using remote sensing data from open sources / A. Kolesnikov, A. Nungesser, P. Kikin. – Текст : непосредственный // E3S Web of Conferences. Серия: Regional Problems of Earth Remote Sensing, RPERS 2020. – 2020: 02006.

30 Using systems of parallel and distributed data processing to build hydrological models based on remote sensing data / A. A. Kolesnikov, P. M. Kikin, E. A. Panidi, A. G. Rusina. – Текст : непосредственный // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives. – 2021. – 24. – P. 111–116.

31 Integration of data of the remote sensing, GIS, and gamma-spectrometric analysis to study soil material redistribution / L. N. Trofimetz, A. A. Kolesnikov, E. A. Panidi, P. M. Kikin. – Текст : непосредственный // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives. – 2021. – 24. – P. 147–152.

32 Kolesnikov, A. Modelling of traffic flows and supply chains based on geospatial knowledge / A. Kolesnikov, P. Kikin, E. Panidi. – Текст : непосредственный // Journal of Physics: Conference Series. Серия: 2021 4th International Conference on Applied Mathematics, Modeling and Simulation, AMMS 2021. – 2021. – 4: 012042.

33 Использование машинного обучения для построения картографических изображений / А. А. Колесников, П. М. Кикин, Е. В. Комиссарова, Д. В. Грищенко. – Текст : непосредственный // От карты прошлого – к карте будущего : сб. научных трудов в 3 т. ; Отв. ред. С. В. Пьянков, 2017. – С. 108–118.

34 Колесников, А. А. Использование открытых семантических баз данных для получения пространственной информации / А. А. Колесников, П. М. Кикин, Е. В. Комиссарова. – Текст : непосредственный // Интерэкспо ГЕО-Сибирь-2017. XIII Междунар. науч. конгр. : Междунар. науч. конф. «Геодезия, геоинформатика, картография, маркшейдерия» : сб. материалов в 2 т. (Новосибирск, 17–21 апреля 2017 г.). – Новосибирск : СГУГиТ, 2017. Т. 2. – С. 53–57.

35 Анализ и обработка данных дистанционного зондирования Земли методами машинного обучения / А. А. Колесников, П. М. Кикин, Е. В. Комиссарова, Д. В. Грищенко. – Текст : непосредственный // Региональные проблемы дистанционного зондирования Земли : сб. материалов V Международной научной кон-

ференции. Сибирский федеральный университет, Институт космических и информационных технологий, 2018. – С. 130–134.

36 Колесников, А. А. Использование четырехмерных сверточных нейронных сетей для автоматизации построения моделей местности / А. А. Колесников, П. М. Кикин. – Текст : непосредственный // Марчуковские научные чтения – 2019 : сб. трудов Международной конференции, 2019. – С. 240–245.

37 Отбор картографических параметров для построения математической модели прогнозирования типов растительного покрова / А. А. Колесников, Е. В. Комиссарова, Т. С. Криволицкая, Д. В. Грищенко. – Текст : непосредственный // Регулирование земельно-имущественных отношений в России: правовое и геопространственное обеспечение, оценка недвижимости, экология, технологические решения. – 2019. – Т. 1. – С. 134–138.