

УТВЕРЖДАЮ

Директор
Федерального государственного
бюджетного учреждения науки
Института динамики геосфер имени
академика М.А. Садовского
Российской академии наук (ИДГ РАН)

 / Турунтаев С.Б.

«5» сентября 2022 г.



ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного бюджетного учреждения науки Института
динамики геосфер имени академика М.А. Садовского Российской академии
наук (ИДГ РАН)

на диссертацию **Малыгина Ивана Вячеславовича**

**«Адаптация алгоритмов машинного обучения к геофизическим задачам
с дефицитом данных»**, представленную на соискание ученой степени
кандидата технических наук по специальности 25.00.10 Геофизика,
геофизические методы поисков полезных ископаемых

Диссертация посвящена исследованию нескольких классических задач с недостатком данных: пространственная интерполяция (двумерная и трехмерная) и классификация на основе временных рядов. Задачи с пропуском данных являются традиционными и **актуальными** для геофизических исследований. Это связано с недостаточным количеством измерений, непродолжительным промежутком времени наблюдений, перерывами в регистрации данных из-за технических неполадок оборудования. Подобная ситуация дефицита данных затрудняет обработку и интерпретацию результатов измерений геофизических полей. Одним из возможных подходов к решению подобных задач является адаптация современных методов машинного обучения к обработке экспериментальных наблюдений.

Новизна работы заключается в ряде факторов. Во-первых, разработан новый метод интерполяции геофизических данных на основе алгоритмов машинного обучения, который применен для построения трёхмерного распределения проводимости среды по данным межскважинного электромагнитного просвечивания. Во-вторых, разработан новый метод расчета трёхмерной сейсмической модели по набору одномерных скоростных разрезов. Применение разработанного метода к данным сейсмических экспериментов SVEKALAPKO и POLENET/LAPNET позволило построить

карту Мохо центральной части Фенноскандии по поперечным сейсмическим волнам и оконтурены области с относительно низкими значениями скоростей S- волн в верхней коре. Показано, что низкоскоростной слой поперечных волн в центральной части Финляндии может быть обусловлен особенностями процесса постледниковой релаксации региона. В-третьих, создана интеллектуальная система краткосрочного прогнозирования ледового заторообразования на основе алгоритма голосования и метода интерполяции неоднородно распределённых гидрометеорологических данных. Эффективность работы интеллектуальной системы продемонстрирована на основе данных с участка р. Северная Двина с оцененной достоверностью прогнозирования 85 %. Интеллектуальная система позволяет получить количественные оценки влияния гидрометеорологических факторов на процесс ледового заторообразования. Показано, что основное влияние оказывает группа гидрологических факторов.

Практическая значимость состоит в создании интеллектуальной системы краткосрочного прогнозирования различных экстремальных природных событий и разработке метода интерполяции неоднородно распределённых геофизических данных. Метод может использоваться для построения пространственных распределений измеряемых величин, например, в геомагнитных исследованиях. Разработанная интеллектуальная система для осуществления краткосрочного прогнозирования мощности процесса заторообразования для участка реки Северная Двина является важной частью прогноза наводнений для данной территории. Разработанная интеллектуальная система может быть применена для данных других регионов, а также для прогнозирования иных опасных природных явлений с аналогичной структурой исходных данных наблюдений.

Диссертационная работа состоит из введения, трех глав, заключения, списка литературы из 168 наименований, четырех приложений. Текст диссертации изложен на 154 страницах машинописного текста, содержит 16 таблиц и 39 рисунков.

Во **введении** обоснована актуальность работы, сформулированы цель и задачи исследования, показаны научная новизна и практическая значимость, представлены защищаемые положения.

В **первой главе** рассмотрены примеры геофизических задач в ситуации недостатка данных, в которых применение современной теории машинного обучения и алгоритмов распознавания образов привело к новым результатам. Введены понятия и определения теории машинного обучения, рассмотрена общая постановка задачи обучения с учителем, приведены основные функционалы качества. В задачах построения 2D- и 3D-моделей среды рассмотрена ситуация недостатка исходных пространственных данных измерений, предложен способ ее решения на основе алгоритма ближайших соседей. В задаче прогнозирования опасных геофизических явлений с ограниченным объемом временных данных предложен способ решения на основе комбинаторно-логического подхода теории распознавания образов. Представлен обзор основных методов решения задач с пропуском данных. Построена трехмерная модель среды при проведении межскважинных

исследований на основе адаптации алгоритмов машинного обучения к данным электромагнитного просвечивания, позволяющая более точно выделить границы слоев по сравнению с методами, используемыми ранее. Проведенные построения обобщены в универсальный метод интерпретации геофизических данных на основе алгоритмов машинного обучения.

В первой главе содержится материал, относящийся к первому защищаемому положению – методу пространственно-временной интерполяции нерегулярно распределённых геофизических данных, основанному на использовании базовых методов машинного обучения, и применимому в задачах с дефицитом данных и сильной анизотропией пространственного распределения измерений. Метод сформулирован и приведен в диссертационной работе в качестве пошаговой последовательности действий для решения задачи интерпретации геофизических данных в условиях их дефицита на основе адаптации алгоритмов машинного обучения.

Во второй главе представлены результаты применения разработанного метода для анализа глубинного скоростного строения региона Фенноскандия на основании данных сейсмических экспериментов POLENET/LAPNET и SVEKALAPKO. С помощью разработанного метода построен ряд карт: уточненная карта границы Мохо на основе набора вертикальных моделей для региона Фенноскандия, карта слоя с низкими значениями скоростей поперечных сейсмических волн V_s в верхней коре изучаемого региона. Показано, что слой низких сейсмических скоростей в верхней коре присутствует на значительной части региона, включая области с протерозойскими породами. Разработанный метод применен к результатам инверсии волновых форм приемных функций для построения пространственной скоростной модели южной части региона. Это позволило детальнее проанализировать структуру верхней части коры под низкоскоростным слоем. Проведенный анализ сейсмических данных показал эффективность разработанного метода на основе адаптации алгоритмов машинного обучения для их анализа и обобщения, что подтверждает первое защищаемое положение.

В третьей главе рассматривается задача построения интеллектуальной системы для краткосрочного прогноза образования заторов в весенний период на северных реках (на примерах рек Северная Двина и Лена) на основе анализа данных коротких временных интервалов с применением специализированных алгоритмов теории машинного обучения (алгоритм вычисления оценок и алгоритм голосования по булевым тестам). Система позволяет проверять гипотезы о влиянии признаков на итоговую мощность явления в условиях ограниченного набора исходных наблюдений на гидропостах и метеостанциях. Построено ранжирование используемых признаков по степени влияния на результат. Проведена оценка достоверности прогнозов на периоде разработки (20 сезонов), и валидация на расширенном периоде (26 сезонов). Оцененная достоверность прогнозирования для этих периодов согласуется и для региона бассейна р. Северная Двина составляет 85%. Разработанная интеллектуальная система применена к другому региону – бассейну р. Лена.

Оцененная достоверность прогнозирования для нового региона составляет 76%. Основные качественные выводы соответствуют полученным ранее результатам.

В третьей главе содержится материал, подтверждающий второе защищаемое положение – разработана и протестирована на реальных данных интеллектуальная система для краткосрочного прогноза образования ледовых заторов в весенний период на северных реках, основанная на анализе данных, полученных по ограниченному набору наблюдений на гидропостах и метеостанциях, с применением специализированных методов теории машинного обучения и распознавания образов.

В **заключении** подводятся итоги работы, делается вывод о том, что в результате проведенных исследований и практических разработок была достигнута цель диссертационного исследования – осуществлена разработка компьютерных систем и методов обработки данных в условиях ограниченного количества данных, недостаточных для проведения классического статистического анализа, на основе методов машинного обучения.

Приведенные в диссертации материалы изложены достаточно подробно, работа снабжена большим количеством иллюстративного материала – схемами, диаграммами, картами, облегчающими восприятие результатов.

Вместе с тем, работа не лишена недочетов:

1. В диссертационном исследовании приведены решения ряда геофизических задач путем адаптации хорошо известных алгоритмов машинного обучения: алгоритма ближайших соседей, алгоритмов голосования и вычисления оценок, алгоритма случайного леса. Такие алгоритмы широко используются в различных областях науки, в том числе и геофизики, тем самым создание интеллектуальной системы на основе таких алгоритмов не может рассматриваться как абсолютно новый результат. При этом, новизна исследований обусловлена, главным образом, не созданием нового метода, а применением метода к решению конкретных геофизических задач.
2. В тексте диссертации при описании решения идентичных задач краткосрочного прогнозирования ледовых заторов с использованием интеллектуальной системы отсутствует объяснение принципов выбора того или иного набора факторов для различных регионов.
3. Для разработанного метода интерполяции пространственно неоднородных данных геофизического мониторинга отсутствуют формальные границы применимости метода в условиях дефицита данных и требования к самим исходным данным.
4. В задаче построения трехмерной модели среды при проведении межскважинного электромагнитного просвечивания (Глава 1) делается вывод, что применение разработанного метода позволяет получить более четкий контрастный образ областей с повышенным коэффициентом затухания по сравнению с использованной ранее интерполяцией на основе линейного кригинга. При этом автором не приводится визуальное сравнение построенных разными методами


образов. Кроме того, отсутствует оценка разрешающей способности разработанного алгоритма интерполяции (не приведен размер трехмерной ячейки выходного grid-файла).

5. В тексте диссертации отсутствует формальное определение понятия «переобучение», что затрудняет обоснование критериев выбора граничного уровня функционала качества, и, как следствие, значений гиперпараметров модели.

Указанные замечания не снижают высокую значимость диссертационного исследования и не ставят под сомнение полученные автором результаты. Диссертация И.В. Малыгина представляет собой законченное исследование, полученные результаты прошли апробацию на многочисленных научных конференциях и семинарах. По теме диссертации опубликовано 7 работ в рецензируемых журналах, входящих в список ВАК РФ, международные реферативные базы данных и системы цитирования Wos и Scopus. На разработанное автором программное обеспечение получено 3 Свидетельства о государственной регистрации программ для ЭВМ. Публикации в полной мере отражают содержание исследования, основные выводы и защищаемые положения диссертации. Текст автореферата соответствует содержанию диссертации.

Диссертационная работа И.В. Малыгина соответствует критериям п. 9 - 11, 13, 14 «Положения о присуждении ученых степеней» (Постановление 6 Правительства РФ от 24.09.2013 г. №842). Иван Вячеславович Малыгин заслуживает присуждения ему ученой степени кандидата технических наук по специальности 25.00.10 Геофизика, геофизические методы поисков полезных ископаемых.

Отзыв составили  01.05.2022 Гоев Андрей Георгиевич,
подпись, дата

и  01.09.2022г. Остапчук Алексей Андреевич.
подпись, дата


Гоев Андрей Георгиевич – кандидат физико-математических наук, старший научный сотрудник лаборатории Сейсмологических методов исследования литосферы Федерального государственного бюджетного учреждения науки Института динамики геосфер имени академика М.А. Садовского Российской академии наук (ИДГ РАН), тел: (495)9397590, e-mail: andr.goev@gmail.com

Я, Гоев Андрей Георгиевич, даю согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

 01.05.2022 Гоев Андрей Георгиевич
подпись, дата

Остапчук Алексей Андреевич – кандидат физико-математических наук, заведующий лабораторией Деформационных процессов в земной коре Федерального государственного бюджетного учреждения науки Института динамики геосфер имени академика М.А. Садовского Российской академии наук (ИДГ РАН), тел: (495)9397511, e-mail: ostapchuk.aa@idg.ras.ru

Я, Остапчук Алексей Андреевич, даю согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

 01.09.2022г.
подпись, дата

Остапчук Алексей Андреевич

Отзыв рассмотрен и обсужден 01.09.2022 на заседании лаборатории сейсмологических методов исследования литосферы Федерального государственного бюджетного учреждения науки Института динамики геосфер имени академика М.А. Садовского Российской академии наук (ИДГ РАН) и одобрен в качестве официального отзыва ведущей организации.

Подписи А.Г. Гоева и А.А. Остапчука заверяю



01.09.2022 Ученый секретарь ИДГ РАН, к.ф.-м.н. Д.Н. Локтев
подпись, дата

Федеральное государственное бюджетное учреждение науки Институт динамики геосфер имени академика М.А. Садовского Российской академии наук (ИДГ РАН)

Адрес: 119334, Ленинский проспект, 38, корпус 1, Москва, Россия

Тел: (499)1376611

E-mail: geospheres@idg.chph.ras.ru

Web: <http://idg.chph.ras.ru>