

## ОТЗЫВ

официального оппонента на диссертацию Малыгина Ивана Вячеславовича «Адаптация алгоритмов машинного обучения к геофизическим задачам с дефицитом данных» на соискание ученой степени кандидата технических наук по специальности 25.00.10 – «Геофизика, геофизические методы поисков полезных ископаемых»

### **Актуальность темы исследования.**

За последние пять-семь лет в мировой практике существенно возросла востребованность машинного обучения и геофизические исследования не являются исключением. Часто эти направления обозначаются более общим термином – искусственный интеллект. Одной из приоритетных задач развития машинного обучения является разработка, апробация и внедрение новых интерпретационных методик и технологий, которые позволят повысить эффективность геофизических исследований. В этой связи диссертационная работа И.В. Малыгина, посвященная исследованию нескольких классических задач с недостатком (дефицитом) данных на основе методов машинного обучения для построения геолого-геофизических моделей и решения задач охраны окружающей среды, является актуальной и своевременной.

Сформулированные диссертантом задачи исследований в полной мере соответствуют поставленной цели работы и обеспечивают ее достижение.

Диссертационная работа состоит из введения, трех глав, заключения, списка литературы из 168 наименований (55% занимают иностранные источники) и четырех приложений, содержит 16 таблиц и 39 рисунков. Текст диссертации изложен на 154 страницах машинописного текста, включая приложения. В приложениях на 18 страницах представлены: таблица данных, используемых для расчетов, структура базы данных прогнозной системы и визуализация ее факторов, список гидропостов и метеостанций в бассейне р. Лена.

**В первой главе** приведен обзор алгоритмов машинного обучения в задачах с дефицитом данных. Диссертантом приводятся факторы дефицита данных, затрудняющие решение различных геофизических задач, которые нейтрализуются путем применения современной теории машинного обучения и алгоритмов распознавания образов. Вводятся понятия и определения теории машинного обучения, рассмотрена общая постановка задачи обучения с учителем, приведены основные функционалы качества и рассмотрены их плюсы и минусы. Представлен наиболее широко распространенный метрический алгоритм распознавания данных – алгоритм ближайших соседей для пространственной интерполяции, и методы создания систем

прогнозирования опасных геофизических явлений. Рассматриваются основные методы обработки в задаче восстановления пропусков данных.

В результате выполненного анализа существующих алгоритмов машинного обучения в задачах дефицита данных диссертантом разработана схема моделирования на основе алгоритмов машинного обучения на примере построения трехмерной модели среды при проведении межскважинного электромагнитного просвечивания. Предложенная схема обобщена в новый метод пространственно-временной интерполяции геофизических данных.

**Во второй главе** приводятся результаты, полученные с помощью разработанного нового метода. Выбрана территория Фенноскандии, для которой у диссертанта имелись все необходимые для расчетов данные – результаты экспериментов POLENET/LAPNET и SVEKALAPKO, при этом все наборы характеризуются дефицитом наблюдений.

Диссертантом рассматривалось три задачи по уточнению проведенных ранее исследований с применением разработанного метода машинного обучения: построение карты поверхности Мохоровичича, построение карты поверхностного низкоскоростного слоя поперечных сейсмических волн и построение трехмерной скоростной модели. В расчетах использовался метод  $k$  ближайших соседей, выполнена оценка средней абсолютной ошибки от числа ближайших соседей  $k$  и подобраны оптимальные значения  $k$  при интерполяции данных.

На основании полученных результатов И.В. Малыгин предлагает новый метод  $k$  построению трехмерной цифровой сейсмической модели по совокупности одномерных зависимостей упругих свойств среды от глубины. Главная сложность решения этой задачи состоит в сильной анизотропии пространственного распределения входных данных, вызванная особенностями расстановки сейсмических станций. Данную проблему предлагается решать масштабным преобразованием горизонтальных координат, при этом отмечаются особенности применимости такого подхода, например, между сейсмическими станциями упругие параметры среды должны плавно изменяться.

Полученные результаты для всех трех задач диссертант сопоставляет с результатами прежних исследований, что позволяет выявить локальные неоднородности, отсутствующие в результатах других исследователей, а в конечном итоге утверждать об эффективности методов машинного обучения для анализа и обобщения данных, имеющих дефицит исходных наблюдений.

**В третьей главе** рассматривается задача построения интеллектуальной системы на основе коротких временных рядов наблюдений для прогноза ледового заторообразования на участках северных рек. В качестве объекта

исследования был выбран регион бассейна р. Северная Двина. Разработанная система без дополнительных доработок исходного кода была применена для региона бассейна р. Лена в аналогичной постановке задачи. Оцененная точность прогнозирования составила от 76% (р. Лена) до 85% (р. Северная Двина).

При реализации данной задачи диссертантом применены методы теории машинного обучения и теории распознавания образов, потребовавшие проведения специальной адаптации под конкретное приложение: относительно исходных алгоритмов была доработана возможность использования непрерывных признаков (в оригинальных алгоритмах использовались только бинарные), предложен способ автоматизированного нахождения множества решающих правил, применена комбинация алгоритмов вычисления оценок и голосования для достижения лучшего качества классификации.

Разработанная И.В. Малыгиным система позволяет проверять гипотезы о влиянии выделенных характерных признаков на итоговую мощность явления в условиях ограниченного набора исходных наблюдений на гидропостах и метеостанциях. Выполнено ранжирование используемых признаков по степени влияния на результат. Полученные результаты согласуются с ранее проведенными исследованиями.

В заключении сделаны основные выводы по диссертационной работе.

Все поставленные в разделах диссертации задачи решены. Вопросы, на которые автор вызвался найти обоснованные ответы, достаточно хорошо проработаны и свидетельствуют о завершенности в целом диссертационной работы. Работы иллюстрированы в достаточном объеме, текст изложен хорошим литературным языком.

**Научная новизна.** В процессе выполнения работы соискателем получены результаты, представляющие научную новизну. На взгляд оппонента, к основным значимым научным результатам исследований можно отнести разработку новых методов: метод построения трехмерного распределения проводимости среды по данным межскважинного электромагнитного просвечивания, метод расчета трехмерной сейсмической модели по набору одномерных скоростных разрезов и метод для анализа многомерных временных рядов ограниченной длины, техническая реализация которого выполнена на основе прогнозной интеллектуальной системы для процесса ледяного заторообразования на северных реках.

**Достоверность, новизна и практическая ценность результатов.** Достоверность диссертации подтверждается взаимосвязанностью результатов разработанных компьютерных систем и методов обработки данных на основе

алгоритмов машинного обучения с существующими геоло-геофизическими моделями. Обоснованность научных положений, выводов и рекомендаций определяется привлечением большого объема фактического материала, современных методов получения и обработки информации, соответствием результатов исследований с ранее опубликованными данными.

Работа построена логично, содержит новые выводы и положения, которые могут быть успешно применены для построения геолого-геофизических моделей и решения задач охраны окружающей среды в условиях ограниченного количества данных, недостаточных для проведения классического статистического анализа.

Практическая ценность результатов исследований заключается в том, что разработанные методы являются достаточно общими, позволяющие интерполировать различные геофизические данные. Разработанная интеллектуальная система может быть применена в любых регионах земного шара, а также использована для прогнозирования различных опасных природных явлений с аналогичной структурой исходных данных о событиях.

В совокупности это позволяет утверждать, что автор диссертационной работы достиг поставленной цели и значимых для практики результатов.

**Соответствие диссертации паспорту научной специальности.** Диссертационная работа И.В. Малыгина соответствует паспорту специальности 25.00.10 – Геофизика, геофизические методы поисков полезных ископаемых по следующим пунктам формулы специальности: методы обработки и интерпретации результатов измерения геофизических полей; интегрированный анализ многомерной, многопараметровой и разнородной информации, включающей геофизические данные; применение геофизических методов при решении задач охраны окружающей среды.

Специальность соответствует технической отрасли наук.

**Обоснованность научных положений, выводов и рекомендаций** подтверждена в 8 научных публикациях по теме диссертации и трех свидетельствах о государственной регистрации программ для ЭВМ. При этом 5 статей опубликовано в изданиях, входящих в международные базы данных WoS и Scopus, и 2 в изданиях, входящих в перечень рецензируемых научных журналов из списка ВАК, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук.

Основные положения, выводы и рекомендации докладывались и обсуждались на международных и всероссийских конференциях.

### **Замечания.**

Представленное диссертационное исследование выполнено на высоком профессиональном уровне, но при этом не лишено некоторых недостатков:

1. На взгляд оппонента в работе недостает демонстрации результата работоспособности методов на искусственно сгенерированной подборке данных без пропусков. Например, можно было бы сгенерировать какое-либо поле с распределенным набором величин. Пусть это поле будет образовано равномерной сеткой, в узлах которых точно известны величины. А потом произвольно убрать часть узлов так, чтобы это выглядело как неравномерная сетка и попытаться восстановить изображение с помощью предложенного метода и сравнить его с тем изображением, что было вначале. Так можно было бы не только оценить сам метод, но и исследовать границы его применимости.

2. Четко не показано какие существуют возможные недостатки у предложенных методов и как их можно устранить, каковы границы применимости методов.

3. Из текста не ясно, почему использование томографии невозможно для построения трехмерной модели среды (стр. 42). В чем заключаются сложности?

4. Недостаточно четко используются некоторые термины. Например, что подразумевается под термином «сильная анизотропия» (стр. 75)? Какой критерий ее силы и что считается слабой анизотропией? Что подразумевается под фразой «мощность возникновения опасного явления» (стр. 83)? В чем измеряется мощность?

5. Отмечается ряд нестыковок, например, как автор может объяснить несоответствие полученных им результатов глубины до Мохо в правой части рис. 2.8, стр. 65, относительно результатов, полученных в работе [Tiira et al., 2014]?

Несмотря на отмеченные замечания, считаем, что работа И.В. Малыгина соответствует уровню кандидатской диссертации и обладает несомненной научной новизной и практической значимостью.

### **Достигнутые результаты диссертационной работы.**

Диссертация является законченным исследованием, выполненным автором самостоятельно. Полученные соискателем результаты достоверны, выводы и защищаемые положения обоснованы. Работа основана на большом фактическом материале и содержит новые выводы и положения, которые могут быть успешно применены для построения геолого-геофизических моделей и решения задач охраны окружающей среды. Автореферат соответствует основному содержанию диссертации.

Диссертационная работа Ивана Вячеславовича Малыгина является научно-квалификационной работой, выполнена на хорошем научном уровне и

удовлетворяет требованиям, установленным Положением о присуждении ученых степеней (утвержденным постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842 в ред. от 20.03.2021, пп. 9, 10, 11, 13, 14), а ее автор, Иван Вячеславович Малыгин, заслуживает присуждения ученой степени кандидата технических наук по специальности 25.00.10 – Геофизика, геофизические методы поисков полезных ископаемых.

Официальный оппонент:

доктор технических наук, заместитель директора по научной работе,  
заведующая лабораторией сейсмологии Федерального государственного  
бюджетного учреждения науки Федерального исследовательского центра  
комплексного изучения Арктики имени академика Н.П. Лавёрова Уральского  
отделения Российской академии наук  
163069, Россия, г. Архангельск, набережная Северной Двины, 23  
e-mail: [essm.ras@gmail.com](mailto:essm.ras@gmail.com)  
телефон: +78182287636

Я, Антоновская Галина Николаевна, даю согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

 АНТОНОВСКАЯ Галина Николаевна

« 1 » сентября 2022 г.

Подпись официального оппонента заверяю:

