

Первый опыт применения технологий машинного обучения для прогноза глубин заложения сейсмического заряда при планировании и проведении полевых сейсморазведочных работ

- Акуленко А.С., ООО НПЦ «Геостра» (АО «Башнефтегеофизика»), г. Уфа
- Гафаров Р.М., АО «Башнефтегеофизика», г. Уфа

В статье представлены результаты исследований в направлении применения технологий машинного обучения и «глубокого обучения» для прогноза параметров возбуждения упругих колебаний (глубин заложения заряда). Рассмотрены ключевые аспекты, связанные с выбором наиболее перспективного алгоритма, подходы к формированию обучающей выборки. На основе производственных данных подтверждена принципиальная возможность решения задачи прогноза параметров данного типа, определены перспективные направления дальнейшего развития методологии работ.

Введение

На сегодняшний день, процессы цифровизации и автоматизации являются ключевыми драйверами повышения экономической эффективности во многих наукоёмких отраслях. Технологии машинного обучения, представляют собой один из наиболее перспективных инструментов реализации данных процессов и активно внедряются в производственные циклы, что во многом обусловлено сформированным на текущий момент колоссальным объёмом разноплановой статистической информации, на основании которой, программные алгоритмы могут устанавливать сложные зависимости одних параметров от других, учитывая множество факторов и аспектов. На основе данных зависимостей, алгоритмы способны принимать определенные решения, например, открытие турникета на проходной, только в случае если программа распознала, что человек находится в защитной одежде или создавать прогностические данные, например прогноз распространения эпидемии в различных странах.

Алгоритмы машинного обучения активно внедряются и в технологические процессы камеральных сейсморазведоч-

ных работ, прежде всего в направлениях синтеза данных ГИС, кинематической и динамической интерпретации. Внедрение технологий машинного обучения в процесс полевых сейсморазведочных работ также имеет значительный потенциал, например для автоматического проектирования смещений физических наблюдений на основе космоснимков или прогноза оптимальных параметров возбуждения упругих колебаний.

Цели и задачи

В связи с тем, что работы по применению технологий машинного обучения, включая «глубокое обучение» для прогноза оптимальных параметров возбуждения упругих колебаний, ранее в компании АО «Башнефтегеофизика» не проводились, и в открытых источниках информация о результатах проведения подобных исследований отсутствовала, основной целью выполненных работ являлось получение подтверждения принципиальной возможности решения задачи в отношении прогноза глубин заложения сейсмического заряда, а также определение ограничений и перспективных путей развития данного направления.

Для достижения указанных целей, посредством проведения серий экспериментов, решены следующие ключевые задачи:

- Определение наиболее перспективного алгоритма машинного обучения;
- Ранжирование параметров, потенциально перспективных для использования в обучающей выборке;
- Определение базовых принципов формирования пространственных зон обучающей выборки.

Исходные данные

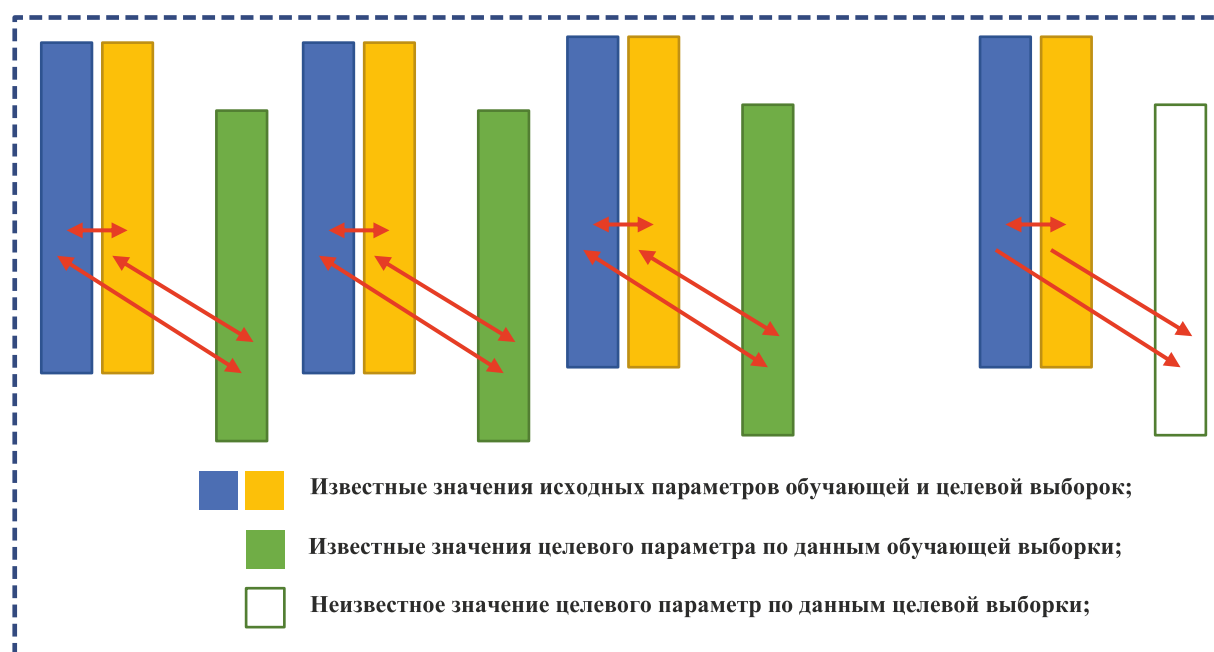
При выполнении работ использован комплекс технологических, динамических и сопутствующих параметров физических наблюдений, отработанных в производственном режиме. Дополнительно привлечена информация о категориях лесных массивов. Указанные данные, на регулярной основе, систематизируются и накапливаются специалистами Департамента разведочной геофизики при выполнении каждого проекта полевых сейсморазведочных работ после внедрения в компании АО «Башнефтегеофизика» системы контроля

качества и мониторинга в 2017 г. [1,2].

При выполнении исследований, в качестве «оптимальных», приняты такие значения глубин заложения заряда, при которых для физических наблюдений обеспечивается величина параметра «доминантная частота» не менее 15HZ и, одновременно, величина «соотношение сигнал/микросейсмы» не менее 10. Значения пороговых критериев определены в соответствии с типовым регламентом приёмки полевых физических наблюдений при выполнении производственных работ.

В процессе выбора априорных параметров, потенциально перспективных для использования при построения прогностических моделей (зависимостей) с помощью алгоритмов машинного обучения, принята во внимание концепция идентичности набора известных параметров (Рис.1), согласно которой известные параметры, которые используются при построении модели, должны быть определены как в массиве обучающей выборки, так и для зон, где целевой параметр необходимо спрогнозировать. При нарушении данного принципа, в

▼ Рис. 1. Схематичное изображение концепции идентичности набора известных параметров.



№П/п	Параметр	Возможность получения параметра до начала регистрации данных
1	Глубина бурения скважин (глубина заложения заряда)	ЦЕЛЕВОЙ ПАРАМЕТР
2	Альтитуда	ДА
3	Соотношение сигнал/микросейсмы	НЕТ
4	Доминантная частота	НЕТ
5	Амплитуда сигнала	НЕТ
6	Амплитуда микросейсм	НЕТ
7	Вертикальное время	НЕТ
8	Вес заряда	НЕТ
9	Приведённая альтитуда	ДА
10	Структурный параметр угла наклона карты альтитуд	ДА
11	Структурный параметр азимута карты альтитуд	ДА
12	Структурный параметр кривизны карты альтитуд	ДА
13	Структурный параметр «шероховатости» карты альтитуд	ДА
14	Литологический состав пород ВЧР (индексация 1-5)	НЕТ
15	Литологический состав пород ВЧР. Бинарный индекс: «Глина»	НЕТ
16	Литологический состав пород ВЧР. Бинарный индекс: «Песок»	НЕТ
17	Литологический состав пород ВЧР. Бинарный индекс: «Плывун»	НЕТ
18	Литологический состав пород ВЧР. Бинарный индекс: «Суглинок»	НЕТ
19	Литологический состав пород ВЧР. Бинарный индекс: «Торф»	НЕТ
20	Категории лесов. Бинарный индекс «Лес1к»	ДА
21	Категории лесов. Бинарный индекс «Лес2к»	ДА
22	Категории лесов. Бинарный индекс «Лес3к»	ДА
23	Категории лесов. Бинарный индекс «Болото»	ДА

▲ Табл. 1. Априорный перечень исходных параметров.

областях отсутствия известных параметров, прогноз будет невозможен или некорректен, в зависимости от используемого алгоритма машинного обучения.

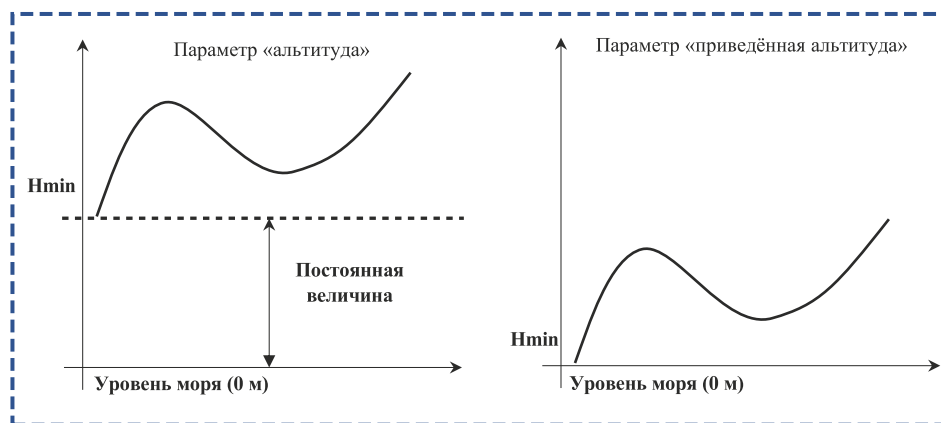
После соответствующей подготовки, включающей редакцию локальных выбросов значений и удаление точек, не удовлетворяющих критериям «оптимальности», для последующих работ были подготовлены двадцать три параметра, перечень которых представлен в Таблице 1, по шести пространственным участкам.

Параметры «литологический состав пород ВЧР» и «категории лесов» рассчи-

таны в бинарном виде, где значение «0» является признаком отсутствия данных конкретного типа, а «1» признаком наличия данных в той или иной пространственной точке.

Параметр «приведённая альтитуда» представляет собой параметр «альтитуда», значения которого изменены посредством вычитания постоянной величины, равной расстоянию от уровня моря до минимального значения параметра (Рис. 2).

С целью определения степени сходимости (наличия устойчивых зависимостей) между целевым параметром «глуби-



▲ Рис. 2. Концепция пересчёта параметра «альтитуда» в параметр «приведённая альтитуда».

на бурения скважин» (глубина заложения сейсмического заряда) и исходными параметрами, рассчитаны соответствующие кросс-плоты как по данным каждого пространственного участка, так и по данным, объединённым в рамках единой пространственной зоны. Анализ зависимостей по данным каждого участка необходим для установления наличия зависимостей параметров на локальном уровне, а анализ по данным после объединения важен для оценки, являются ли зависимости унаследованными, то есть характерными для всех пространственных участков одновременно.

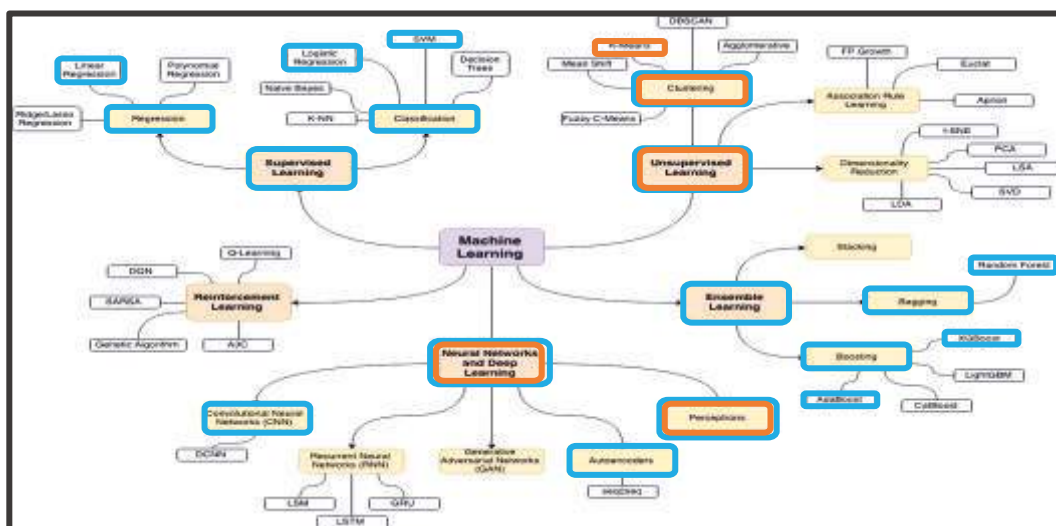
Анализ значений коэффициентов и облик кросс-плотов показал отсутствие устойчивых и унаследованных зависимостей по всем рассматриваемым параметрам. Данный эффект, в целом, являл-

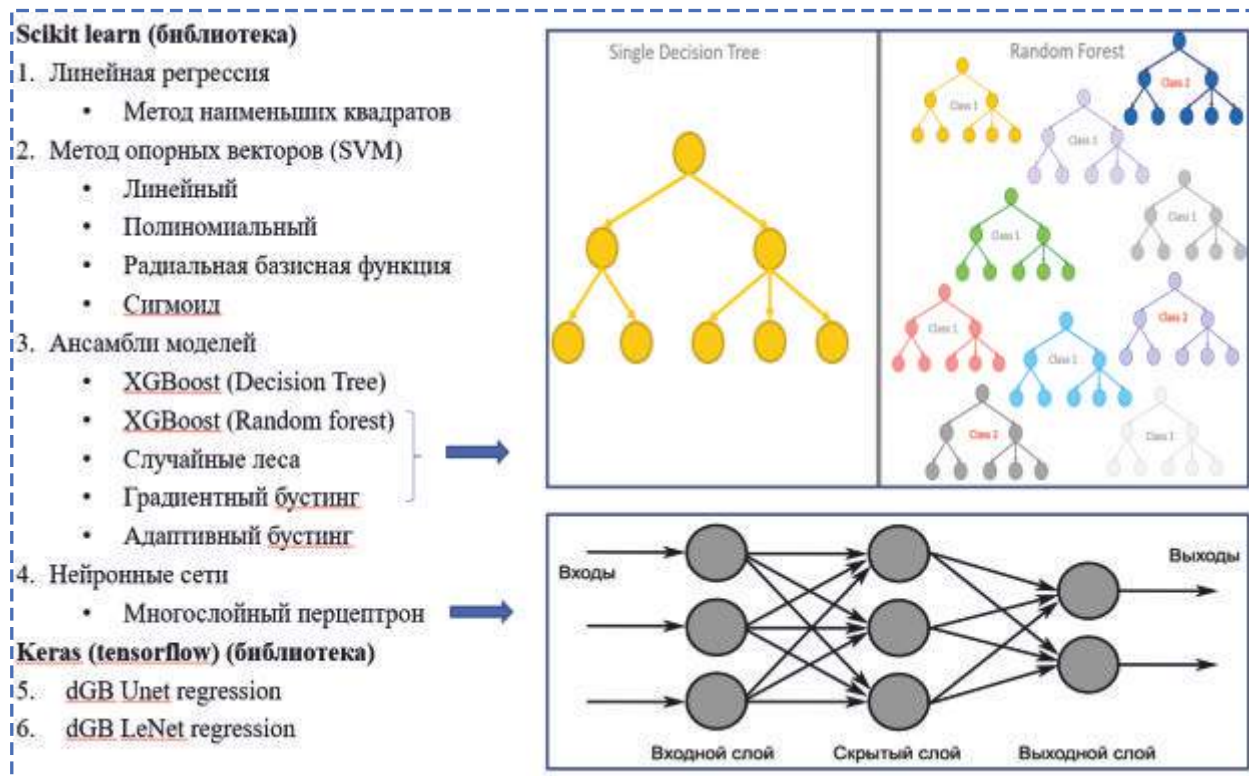
ся ожидаемым, т.к. определение необходимого и достаточного значения глубины заложения заряда в процессе производственных работ определяется на основе комплекса критериев, но в то же время отсутствие стабильных априорных зависимостей значительно усложняет вопрос выбора исходных параметров для формирования обучающей выборки и анализ получаемых результатов.

Определение наиболее перспективного алгоритма машинного обучения

В настоящее время, количество разработанных алгоритмов машинного обучения исчисляется десятками (Рис.3). Математический базис данных алгоритмов, во многих случаях, находится в свободном доступе, так как сами по себе алгоритмы представляют гораздо мень-

▼ Рис. 3. Пример классификации алгоритмов машинного обучения.





▲ Рис. 4. Перечень тестируемых алгоритмов машинного обучения.

шую ценность, чем модели, созданные на их основе и позволяющие решать ту или иную задачу с требуемой точностью.

Большинство алгоритмов машинного обучения, априорно не имеет специализации в отношении решения той или иной задачи, поэтому процесс их тестирования представляет собой неотъемлемую часть выполнения работ при решении задач, отраслевой опыт реализации которых не значителен или отсутствует.

В рамках выполненных исследований, априорный перечень алгоритмов был определен по критерию оперативной доступности и являлся достаточно представительным (Рис. 4), так как включал в себя не только классические алгоритмы машинного обучения, но и «глубокого обучения» (Deep Learning).

Исходя из принципов работы алгоритмов, ожидалось получение наиболее эффективных решений при построении моделей с помощью классов «Ансамбли моделей» (Ensemble models) и «Нейронные сети» (Neural Network), но для того,

чтобы осуществить объективный выбор, тестирование выполнено для полного перечня алгоритмов.

В рамках серии экспериментов, проводимых для решения задачи выбора наиболее перспективного алгоритма, использовались различные наборы параметров обучающей выборки. Данный подход позволил имитировать различные сценарии работы и оценить степень устойчивости алгоритмов к таким ситуациям как: наличие в обучающей выборке параметров «параллельного переноса», то есть параметров, значения которых крайне схожи на участке прогноза и участках, выбранных для обучения, но при этом значения целевого параметра, которые им соответствуют, являются различными; наличие в обучающей выборке «противоречивых данных», то есть наличие большого количества параметров взаимная сходимость значений которых крайне низкая и т.д.

По итогам реализации более тридцати разноплановых экспериментов, в качес-

тве наиболее перспективного, определённого алгоритм «Random forest». Данный алгоритм характеризуется наиболее стабильными результатами реализации при наличии в обучающей выборке данных «параллельного переноса», высокой устойчивостью к наличию «противоречивых данных», отсутствием аномальных локальных значений после реализации рассчитанной модели.

Важно отметить, что результаты, полученные с помощью алгоритма «Многослойный перцептрон» (Multi-Layer Perceptron), который рассматривался перед выполнением тестирования как один из наиболее перспективных, оказались в значительной степени нестационарными, то есть зависимы от внутренней тестовой выборки, которую использует любой алгоритм для валидации модели и принятия решения о прекращении обучения. Данная выборка случайным образом формируется из общей обучающей выборки (15-20% от общего объёма точек), предоставляемой алгоритму для обучения, и является разной при каждом запуске. Таким образом, при условии запуска одного и того же алгоритма с одними и теми же параметрами

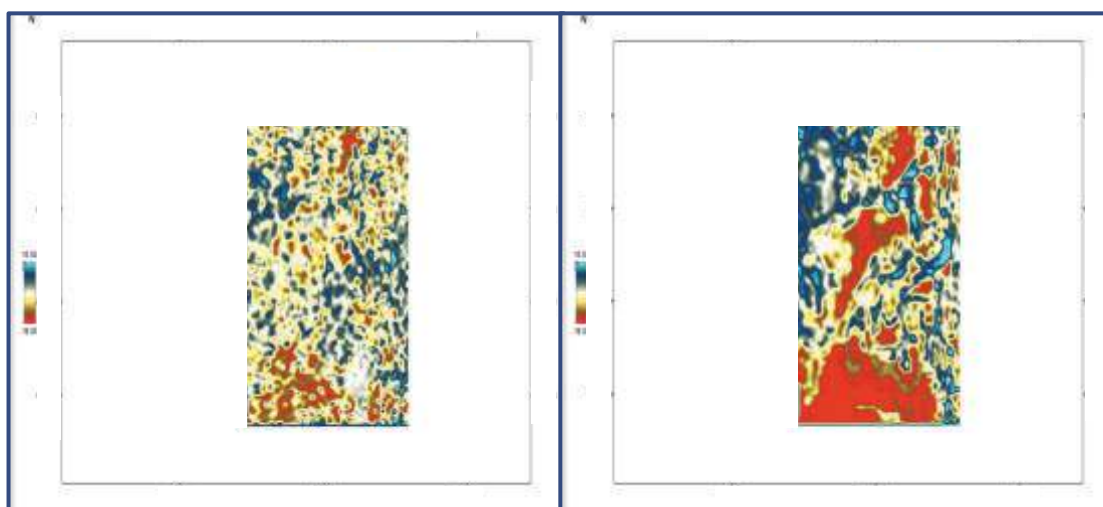
и массивом данных для обучения, результирующая модель и, как следствие, результаты прогноза целевого параметра, будут отличаться.

В процессе тестирования установлено, что для алгоритмов «глубокого обучения» различие результатов реализации моделей не является значимым и составляет не более 2%, тогда как для алгоритма «Многослойный перцептрон» расхождение результатов, полученных при прочих равных условиях, составило от 1% до 20%. Таким образом, несмотря на априорную перспективность, дальнейшее применение алгоритма «Многослойный перцептрон» класса «Нейронные сети» для решения текущей задачи было признано нецелесообразным.

Дополнительно, в процессе выполнения тестирования алгоритмов, был проведен анализ целесообразности сглаживания параметров перед их использованием в обучающей выборке.

Эксперименты показали, что пространственное сглаживание, даже на небольших базах, позволяет значительно снизить осцилляцию результатов прогноза (Рис.5). По данной причине при выполнении дальнейших экспериментов, апри-

▼ Рис. 5. Пример результатов реализации алгоритма «Random forest» с параметрами обучающей выборки без предварительного сглаживания (слева) и после предварительного сглаживания (справа).



орные параметры были сглажены, за исключением значений целевого параметра и параметров, представленных в бинарном виде.

Ранжирование параметров, потенциально перспективных для использования в обучающей выборке

Принимая во внимание опыт, полученный в рамках экспериментов предшествующего этапа, а также учитывая физический смысл параметров, было проведено их ранжирование (Таблица 2) по критерию перспективности использования в обучающей выборке.

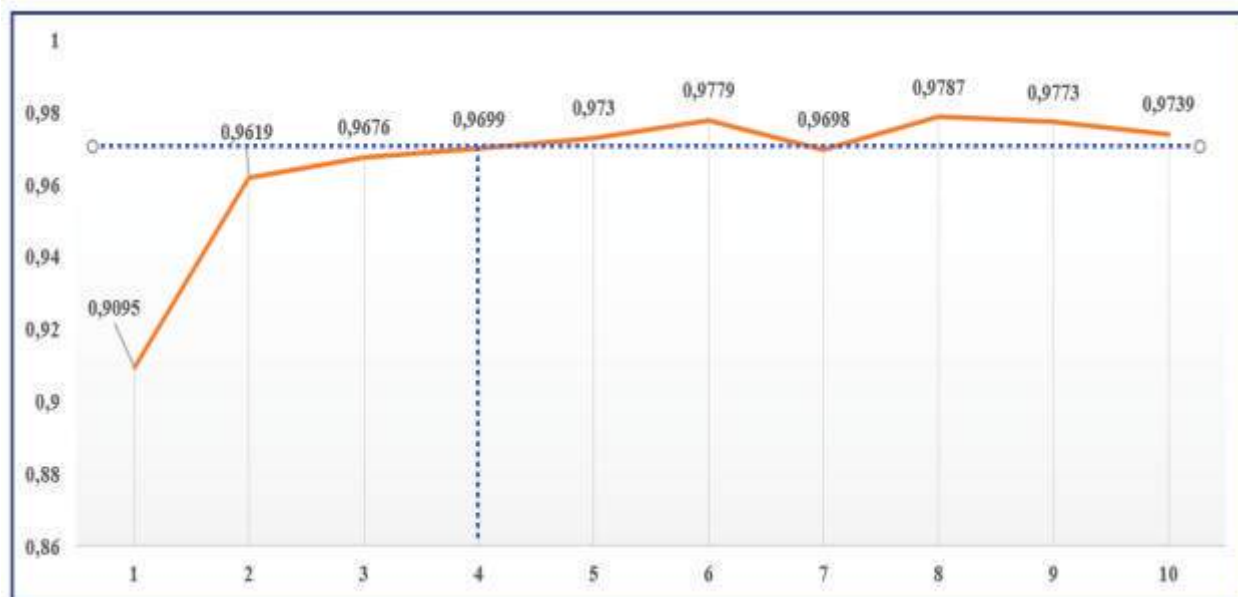
Наибольший приоритет отдан параметрам, которые в той или иной степени описывают литологические и упругие

свойства ВЧР. Исключение составляет параметр «приведённая альтитуда» и группа структурных параметров (геометрические свойства поверхности альтитуд). Данные параметры, как и группа параметров «категория лесов», так же включены в ранжированный перечень, так как они могут быть получены до начала регистрации производственных данных. Некоторые параметры, по различным причинам, были исключены из перечня, например:

- Использование в дальнейших экспериментах параметра «альтитуда» признано нецелесообразным, так как значения имеют постоянную составляющую в виде расстояния от уровня моря до минимального значения параметра.

▼ Табл. 2. Перечень исходных параметров, ранжированных по критерию перспективности использования в обучающей выборке.

№П/п	Параметр	Ранг
1	Доминантная частота	I Параметры являются критериями оптимальности при выборе значений упругих колебаний в процессе производственных работ и косвенно описывают упругие свойства верхней части разреза
2	Амплитуда сигнала	
3	Литологический состав пород ВЧР. Бинарный индекс: «Глина»	II Параметры напрямую или косвенно описывают литологические свойства верхней части разреза
4	Литологический состав пород ВЧР. Бинарный индекс: «Песок»	
5	Литологический состав пород ВЧР. Бинарный индекс: «Плывун»	
6	Литологический состав пород ВЧР. Бинарный индекс: «Суглинок»	
7	Литологический состав пород ВЧР. Бинарный индекс: «Торф»	
8	Вертикальное время	
9	Приведённая альтитуда	III «Структурные» параметры, которые могут быть получены до начала производственных работ
10	Структурный параметр угла наклона карты альтитуд	
11	Структурный параметр азимута карты альтитуд	
12	Структурный параметр кривизны карты альтитуд	
13	Структурный параметр «шероховатости» карты альтитуд	
14	Категории лесов. Бинарный индекс «Лес1к»	
15	Категории лесов. Бинарный индекс «Лес2к»	
16	Категории лесов. Бинарный индекс «Лес3к»	
17	Категории лесов. Бинарный индекс «Болото»	



▲ Рис. 6. Гистограмма изменения значений коэффициента валидации от количества используемых параметров обучающей выборки.

Наличие данной составляющей значительно ограничивает зоны пространственных границ обучающей выборки и усиливает фактор «противоречивости данных». Для решения целевой задачи в большей степени важным является степень пространственного изменения значений превышения рельефа, что в полной мере отражается в параметре «приведённая альтиту́да».

- Значения параметра «амплитуда микросейсм» и, как следствие, параметра «соотношение сигнал/микросейсм», во многих случаях отражают различные внешние (наземные) факторы, которые не имеют причинно-следственных связей с целевым параметром. Перспективы использования данных параметров являются спорными, в связи со значительными рисками искажения модели. По данной причине, в процессе последующих работ указанные параметры не рассматривались.

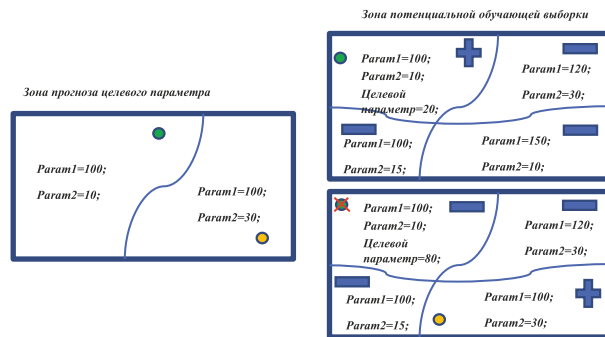
С целью определения минимально необходимого количества параметров обучающей выборки, которое обеспечивает построение достаточно сложной модели для получения стабильных

результатов её реализации, проведены эксперименты с последовательным увеличением количества параметров и анализом коэффициентов валидации модели.

На основе гистограммы (Рис.6), отражающей изменение значений коэффициента валидации от количества используемых параметров обучающей выборки, а также комплексного анализа результатов прогноза, установлено, что минимальное количество параметров обучающей выборки должно составлять не менее четырёх. Данное количество параметров использовано при выполнении последующих экспериментов.

Определение базовых принципов формирования пространственных зон обучающей выборки

Определение пространственных зон обучающей выборки является одной из наиболее важных и, в то же время, сложных стадий решения целевой задачи. Учитывая отсутствие устойчивых исходных взаимосвязей целевого параметра и параметров ранжированного перечня, принимая во внимание, что фактические



▲ Рис. 8. Схематичное изображение реализации принципа непротиворечивости известных значений целевых параметров обучающей выборки.

Несмотря на относительную простоту сформулированных принципов, их комплексная реализация становится крайне сложной задачей при условии работы с большими массивами пространственных участков.

Python-скрипты, разработанные в АО «Башнефтегеофизика» и применяемые для автоматического определения наиболее перспективных элементов обучающей выборки при выполнении работ в направлении синтеза данных ГИС, для решения текущей задачи оказались не эффективны. Требовалась значительная модернизация скриптов, которая бы позволила автоматизировать решение следующих задач:

- Кластеризация целевой пространственной зоны по критериям гомогенности той или иной комбинации параметров, на основе которых планируется выполнять прогноз с возможностью настройки и управления количеством кластеров;

- Поиск зон на всех доступных пространственных участках, в максимально высокой степени сходимых, при выбранной комбинации параметров, с тем или иным кластером целевой пространственной зоны, с возможностью настройки критерия выбора зоны в зависимости от указанного допустимого процента сред-

неквадратичного интегрального расхождения параметров;

- Проверка на отсутствие противоречий в значениях целевого параметра, осуществляемая на основе допустимой величины и вторичная селекция зон по данному критерию.

Учитывая необходимость дополнительной адаптации для работы со значительными (сотни тысяч точек анализа для каждого параметра) объёмами данных, разработка скрипта, который позволил бы автоматизировать процесс определения пространственных границ обучающей выборки, однозначно является отдельной трудоёмкой задачей, решение которой было запланировано в рамках отдельных работ.

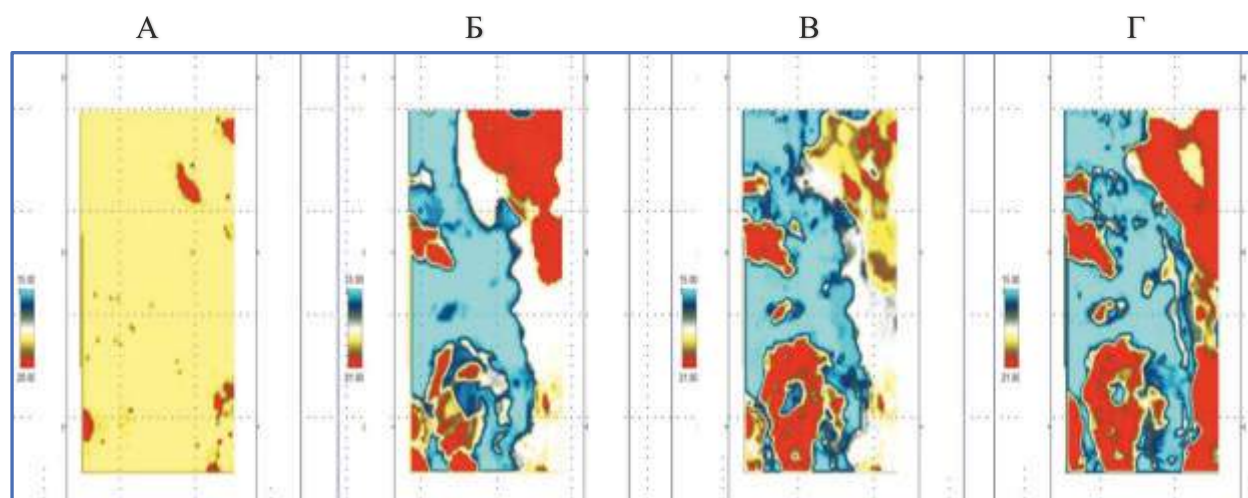
Реализация базовых принципов

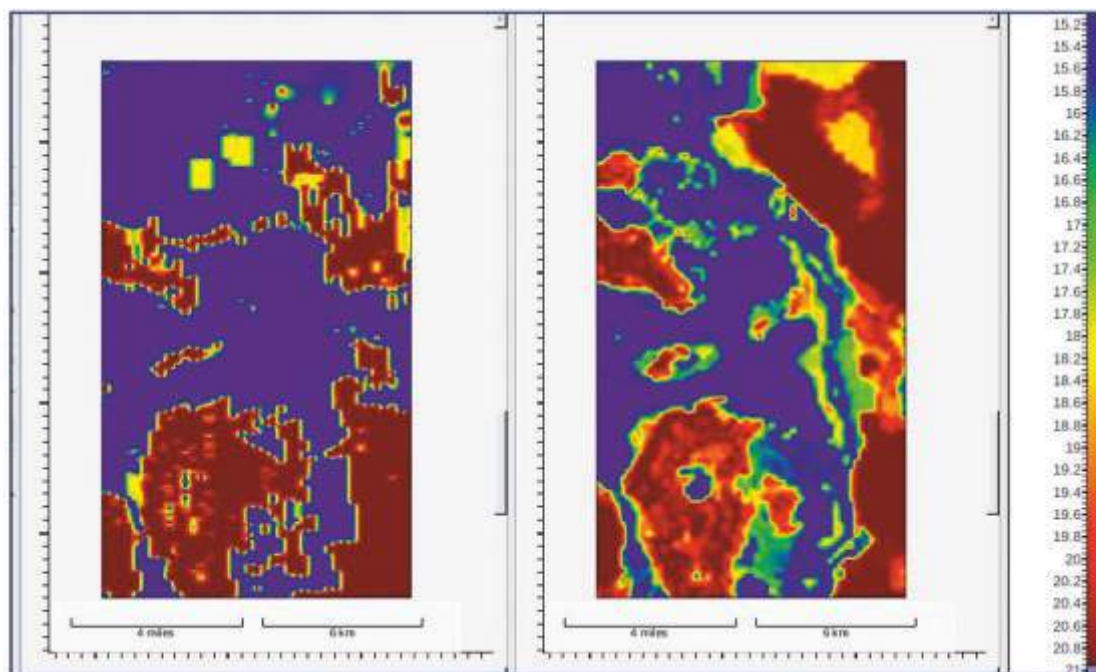
На заключительном этапе работ, консолидировав накопленный в рамках предшествующих экспериментов опыт, с использованием материалов всех шести загруженных пространственных участков, был выполнен цикл расчётов, направленный на получение результатов прогноза при имитации различных

стадий работ, а именно: на стадии «подготовки (планирования) к выполнению производственных работ» – в данном случае обучающая выборка формировалась исключительно из данных участков, не являющихся целевыми; на стадии «производственных работ», при условии отработки 10% и 30% целевого участка – обучающая выборка формировалась как из данных целевого так и иных участков.

Перед выполнением расчётов, на основе сформированных принципов, был проведен анализ зон, перспективных для включения в обучающую выборку на основе кластеризации целевого участка. К сожалению, параметры целевого участка оказались достаточно уникальными и только один из пяти оставшихся участков, частично удовлетворял необходимым условиям. Значения целевого параметра в пределах данного участка являлись постоянными и составляли 18 метров. Как следствие, результат прогноза (Рис.9, а) так же соответствовал этой величине, в то время как фактические значения прогнозируемого параметра на целевой площади составляли 15, 21 метр.

▼ Рис. 9. Результат прогноза глубины заложения заряда при различных способах формирования обучающей выборки: а – использованы данные стороннего участка; б - использованы данные стороннего участка и 10% данных целевого участка; в - использованы данные стороннего участка и 30% данных целевого участка; г – использованы данные целевого участка (30%);





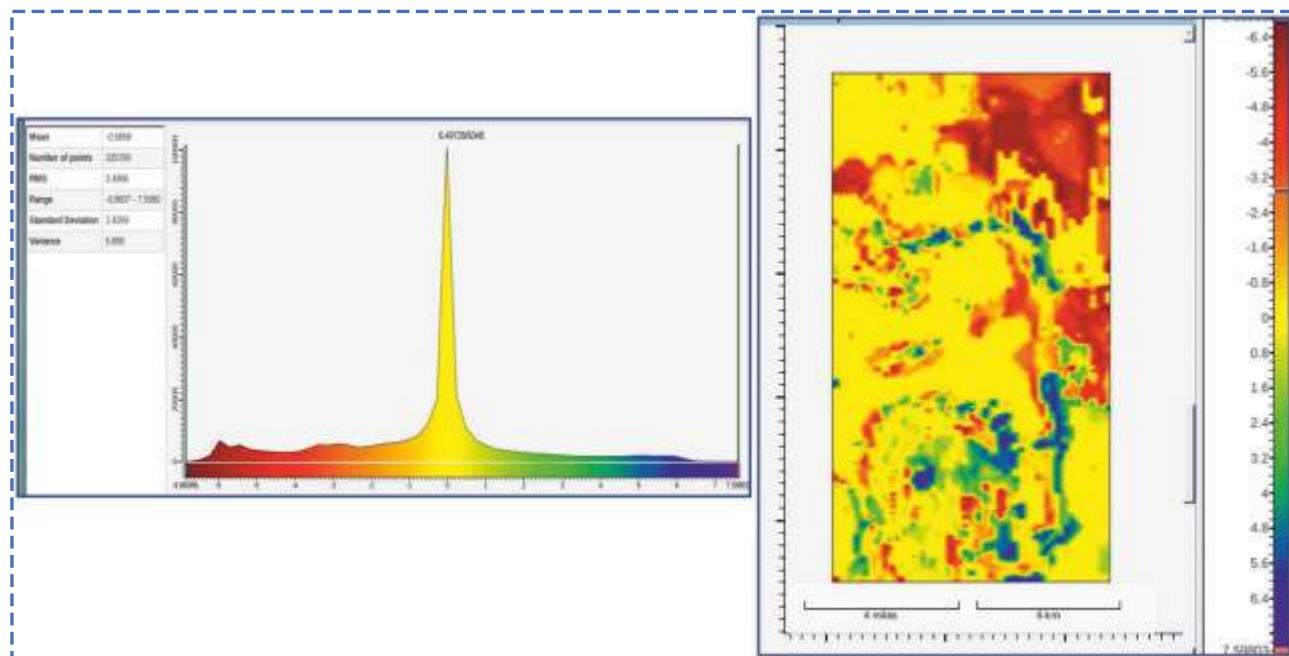
▲ Рис. 10. Сравнение фактических и прогнозных глубин.

Полученный результат, с одной стороны, подтверждает возможность значительного расхождения прогнозных и фактических значений целевого параметра на начальной стадии работ, то есть до момента, когда будут отработаны первые физические наблюдения и полученный диапазон значений целевого параметра позволит провести более корректный выбор зон обучающей выборки. С другой стороны, данный результат, позволяет сделать однозначное заключение о целесообразности подготовки и загрузки полного объема накопленной информации по всем пространственным участкам, отработанным ранее для того, чтобы иметь возможность увеличить количество зон потенциально перспективных для включения в обучающую выборку и, как следствие, повысить точность прогноза на данной стадии.

Привлечение в качестве обучающей выборки не только данных стороннего участка, но и всего 10% данных целевого, позволило получить результаты прогноза, которые в большей мере соответство-

вали фактическим значениям целевого параметра (Рис.9, б). Данная тенденция логичным образом сохранилась и при дальнейшем увеличении процента (30%) данных целевого участка (Рис.9, в), но на северо-востоке, в связи с высокой схожимостью исходных параметров целевого и стороннего участков, прогнозируемые значения имели значительные отклонения от фактических, что было компенсировано при формировании обучающей выборки исключительно из 30% данных целевого участка (Рис.9, г), после чего полученный результат был сопоставлен с фактическими глубинами заложения заряда, полученными при отработке площади (Рис.10). Дополнительно были рассчитаны отклонения фактических и прогнозных глубин в метрах, (Рис.11).

Анализ отклонений показал, что около 20% прогнозных глубин имеют отклонения от фактических более чем на 1,5 метра (длина бурового шнека), что говорит о том, что полученные данные не полностью удовлетворяют необходимым критериям точности, но в то же время неоспоримым положительным



▲ Рис. 11. Гистограмма и карта распределения значений отклонения величин прогнозированных значений целевого параметра от фактических, в метрах.

аспектом является тот факт, что разница суммарных глубин фактических и прогнозных скважин составляет менее 2%, то есть прогнозируемые результаты можно использовать для планирования объема буровых работ (количество бурстанков, персонала, техники и т.д.).

Заключение

Первый опыт применения технологий машинного обучения для прогноза глубин заложения сейсмического заряда, подтвердил принципиальную возможность решения целевой задачи, но, в то же время, выполненный комплекс исследований показал, что данное решение требует значительного объема аналитических работ и представляет собой многостадийный и нетривиальный процесс.

В рамках работ выполнено более ста различных экспериментов, по итогам анализа результатов которых, осуществлен выбор наиболее перспективного алгоритма машинного обучения, проработаны и определены основополагаю-

щие принципы выполнения таких ключевых этапов, как определение априорных наборов параметров, перспективных для включения в обучающую выборку, формирование пространственных зон обучающей выборки.

Несмотря на то, что фактические значения целевого параметра, полученные по отработанным ранее пространственным участкам, не являются эталоном и вполне возможно, что прогнозные значения, наоборот, могут быть более оптимальны, достигнутые на данный момент результаты, не полностью удовлетворяют требованиям точности предсказаний по глубине конкретной скважины, но позволяют более обоснованно планировать объем буровзрывных работ.

Продолжение исследований признано целесообразным в следующих направлениях:

- Подготовка и загрузка полного объема накопленной информации по всем пространственным участкам, отработанным ранее;

- Разработка адаптированного Python–скрипта для автоматизации определения пространственных зон, перспективных для включения в обучающую выборку;
- Верификация получаемых результатов прогноза непосредственно в процессе выполнения полевых сейсморазведочных работ.

Литература

1. Ахметшин И.Н. Сейсморазведка. Испытание временем / И.Н. Ахметшин, Р.М. Гафаров, И.А. Сираев // Геофизика. - 2017. - №3. - С.9-13.
2. Система контроля качества полевых сейсморазведочных работ / Р.М. Гафаров [и др.] // Геология, геофизика и разработка нефтяных и газовых месторождений. - 2021. - № 6 (354). - С. 34-41.

Подробнее об авторах



Акуленко
Александр Семёнович

Главный геофизик ООО
НПЦ «Геостра»
(АО «Башнефтегеофизика»)
Окончил Гомельский ГТУ
имени П.О. Сухого в 2007 г.
Научные интересы –
– полевые и камеральные
сейсморазведочные
работы.



Гафаров
Радий Марсович

Главный геофизик
Дирекции разведочной
геофизики
АО «Башнефтегеофизика».
Окончил Свердловский
Горный Институт в 1985г.
Научные интересы –
– поиск и разведка
нефтяных и газовых
месторождений.