



НОВЫЕ ПОДХОДЫ В АНАЛИЗЕ ГЕОЛОГО-ГЕОФИЗИЧЕСКОЙ ИНФОРМАЦИИ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

**С.А. Буденный, А.Ю. Бухарев, Н.А. Волков, А.П. Цанда (ООО «Инжиниринговый центр МФТИ»),
А.С. Маргарит, А.М. Андрианова, Е.А. Жуковская, Б.В. Белозеров (ООО «Газпромнефть НТЦ»)**

Рассмотрены примеры применения современных подходов на базе алгоритмов машинного обучения для анализа геолого-геофизической информации. Применительно к задаче анализа качества данных о свойствах пластовых флюидов (PVT-свойств) предложен статистический подход, основанный на смеси многомерных распределений Стьюдента. Реализованный подход позволяет выявлять потенциально аномальную пробу (свойства пробы сильно отличаются от существующих аналогов), после чего предлагает значение по выборке существующих аналогов. В задаче интерпретации данных скважинных каротажей предложен подход для верификации геофизического исследования скважин с применением независимой экспертной системы, обученной на результатах ранее интерпретированных скважинных исследований. В частности, при неоднозначности определения коллектора в заданном геологическом разрезе по данным геофизических исследований скважин (ГИС) предлагается использовать такую независимую экспертную систему. Другим принципиально новым подходом к автоматической обработке геолого-геофизической информации является обработка цифровых снимков изображений шлифов терригенных пород с целью автоматической сегментации структурных объектов с применением гибридного подхода, объединяющего спектральные методы обработки изображений и сверточные нейронные сети. Такой подход позволяет оперативно (обработка одного изображения занимает 1...60 с расчетного времени) получать, в частности, морфологические данные обломочных зерен терригенных пород, что позволяет освободить экспертов от рутинной работы, сфокусировать внимание на решении более нетривиальных задач, требующих уникальных экспертных знаний.

Ключевые слова: PVT, геофизические исследования скважин, шлифы, большие данные, машинное обучение, анализ данных, экспертная система.

Введение

На сегодняшний день рентабельное вовлечение запасов углеводородов неизбежно требует либо переосмотра существующих технологий добычи, либо создания принципиально новых подходов. Не являются исключением и залежи, относящиеся к категории трудноизвлекаемых. В то время как недропользователи все больше осознают необходимость адаптации существующих технологий под особенности трудноизвлекаемых запасов, например, перехода на тандем технологий горизонтального бурения и многостадийного гидроразрыва пласта, открытым остается вопрос о способах внедрения цифровых технологий, позволяющих косвенно или напрямую повысить рентабельность освоения и разработки залежей.

Ввиду того, что в период высоких цен на нефть ИТ-отрасль, имеющая дальние горизонты окупаемости, не рассматривалась как драйвер развития нефтегазовой промышленности, на сегодня среди компаний-операторов наблюдается так называемая логистическая инерционность: логистика данных не подготовлена для масштабного внедрения современных методов обработки больших данных (в том числе на базе методов машинного обучения) в отличие от других отраслей, таких как банковский сектор, здравоохранение, телекоммуникации и т.д. Несмотря на это, нельзя отрицать наличие первых плодов внедрения методов

big data в нефтегазовой отрасли. В частности, одной из уже успешно решаемых задач является увеличение скорости и точности обработки промысловых данных. В то же время отсутствие единой принятой структуры в представлении и хранении промысловых данных по добыче, их большой объем и низкая достоверность затрудняют оперативное принятие решений.

ООО «Инжиниринговый центр МФТИ» совместно с ООО «Газпромнефть НТЦ» развивает ряд подходов на базе современных алгоритмов анализа данных для повышения качества промысловых данных. Примерами решенных задач являются поиск аномальных значений в PVT-свойствах (Pressure, Volume, Temperature) пластовых флюидов, поиск продуктивных нефтенасыщенных слоев в геологическом разрезе по данным скважинных каротажей, автоматизация петрографического анализа цифровых изображений шлифов терригенных пород. Рассмотрим решения представленных выше задач на реальных геолого-геофизических данных.

Метод определения аномальности пробы пластовых флюидов

Обоснование компонентного состава и PVT-свойств пластовых флюидов является одним из важнейших условий повышения достоверности подсчета запасов и эффективности проектирования разработки

месторождений. Существует ряд методов оценки представительности проб пластовых флюидов, таких как проверка герметичности пробоотборных камер, сопоставление давления насыщения нефти с давлением сепарации при температуре сепарации и др.; метод Хоффмана-Крампа-Хоккота, основанный на корреляции констант равновесия; определение представительности проб по критерию загрязненности технологическими жидкостями, применяемыми при бурении, перфорации и освоении скважины. При наличии же только сырых данных выше представленные методы не могут быть применены, в связи с чем возникает необходимость в разработке алгоритмов выявления потенциально некорректных значений по сырым данным.

Для практического применения исследований PVT-свойств флюидов была проанализирована корпоративная база данных, содержащая результаты исследований более 2000 проб пластовых флюидов. Среди рассматриваемых параметров имеются пластовое давление, температура пласта, газосодержание и др. Разработана описывающая данные вероятностная модель, с помощью которой производится кластеризация флюидов на конечное число групп, а также производится оценка степени доверия к результатам исследований.

Предложенный метод заключается в описании имеющихся данных с помощью вероятностной модели смеси многомерных распределений Стьюдента. В такой модели плотность данных имеет вид $p(x) = \sum_{j=1}^K w_j T(x|\mu_j, \Sigma_j, \nu)$, где $p(x) = \sum_{j=1}^K w_j T(x|\mu_j, \Sigma_j, \nu)$, $T(x|\mu_j, \Sigma_j, \nu)$ — плотность многомерного распределения Стьюдента (t-distribution). В данной модели име-

ется K кластеров с центрами в точках μ_j и некоторыми параметрами разброса Σ_j . Веса кластеров задаются параметрами w_j .

Оценка параметров модели производится при помощи вариационного байесовского вывода (variational bayesian inference) [1]. Результат оценки параметров для трех кластеров приведен на рис. 1 слева. В качестве более простой модели можно рассмотреть подобную смесь распределений, в которой вместо распределений Стьюдента используются многомерные нормальные распределения. Параметры такой модели оцениваются с помощью EM-алгоритма (Expectation-Maximization).

Результат оценки параметров для смеси из четырех компонент приведен на рис. 1 справа. Можно заметить, что серый кластер накрывает два других кластера, что обусловлено наличием в данных шумовых объектов, для детектирования которых строится описываемый метод. Кластеры, полученные с помощью модели смеси многомерных распределений Стьюдента, соответствуют “нешумовым” кластерам в модели смеси многомерных нормальных распределений, поэтому для решения поставленной задачи более оправдано использовать модель смеси многомерных распределений Стьюдента.

Кластеризация проб происходит по следующему принципу — проба x относится к тому кластеру, которому соответствует максимальная из величин $w_j T(x|\mu_j, \Sigma_j, \nu)$.

Плотности многомерного распределения Стьюдента $T(x|\mu_j, \Sigma_j, \nu)$ вычисляются с помощью семплирования методом Монте-Карло.

Проба считается аномальной, если в соответствующей точке плотность $p(x)$ меньше некоторого

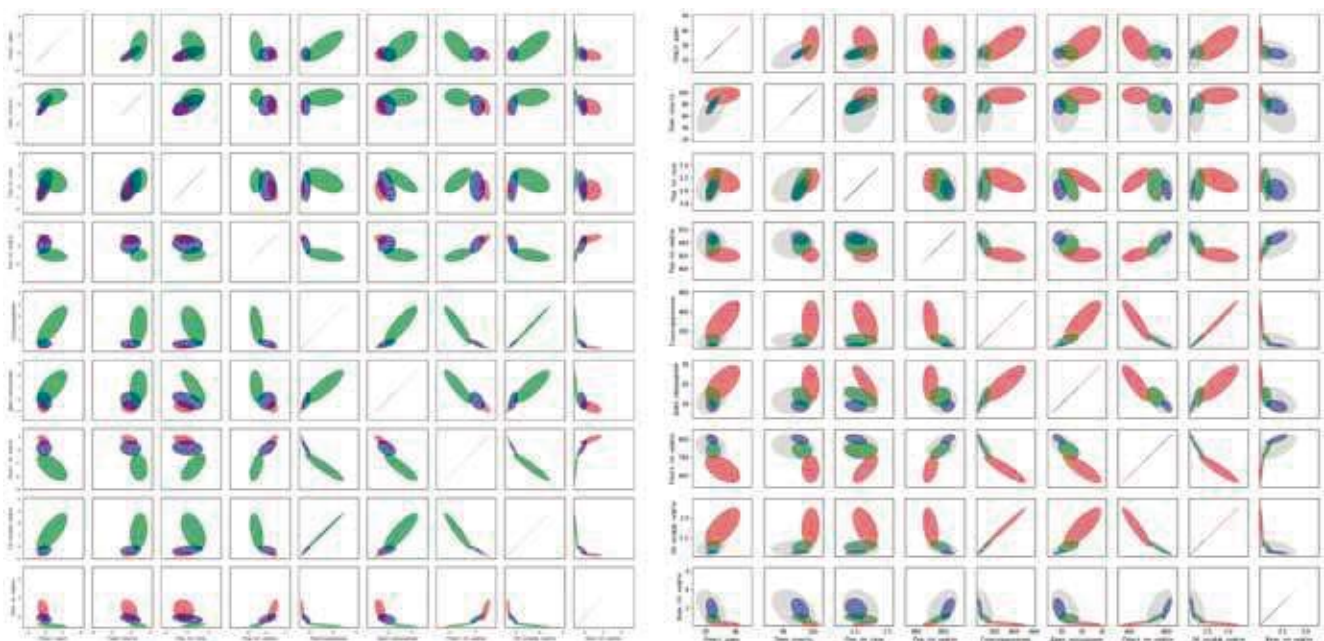


Рис. 1. Визуализация кластеризации с помощью модели смеси многомерных распределений Стьюдента (слева) и многомерных нормальных распределений (справа) в проекциях на двумерные плоскости. Эллипсы соответствуют наиболее типичным значениям каждого кластера. Цвет эллипса соответствует кластеру

порогового значения q . Величина q выбирается как значение плотности $p(x)$, при котором вероятность получить пробу с плотностью, не превосходящей q , в точности равна 0,05. Описанная процедура определения аномальных проб является частным случаем процедуры проверки статистических гипотез.

Если проба была признана аномальной, исследователю предоставляется возможность указать параметры пробы, которым стоит доверять. Для всех остальных параметров реализована процедура расчета наиболее вероятных значений, которые соответствуют доверяемому в рамках определенного ранее кластера пробы.

Все описанные методы работают только в случае, если в пробе известны все значения, то есть отсутствуют пропуски. При наличии в пробе пропусков они предварительно заполняются наиболее вероятными значениями с учетом заполненных значений. Иначе говоря, рассчитывается условное математическое ожидание пропущенных значений при условии известных.

Интерфейс инструмента на базе описанного подхода приведен на рис. 2. Введенная проба является аномальной, поскольку значение ее вероятностной плотности меньше порогового значения. Первые четыре параметра выбраны как доверяемые, по ним производится расчет наиболее вероятных значений остальных параметров.

50.6	NaN	<input checked="" type="checkbox"/>	Пластовое давление, Мпа
86	NaN	<input checked="" type="checkbox"/>	Температура пласта, °C
1.086	NaN	<input checked="" type="checkbox"/>	Стандартная сепарация - Поверхностная плотность газа, кг/м3
830	NaN	<input checked="" type="checkbox"/>	Стандартная сепарация - Поверхностная плотность нефти, кг/м3
240	NaN	<input type="checkbox"/>	Стандартная сепарация - Газосодержание, м3/т
Предлагаемое значение 505.81			
22.8	NaN	<input type="checkbox"/>	Давление насыщения, Мпа
Предлагаемое значение 27.54			
578	NaN	<input type="checkbox"/>	Стандартная сепарация - Пластовая плотность нефти, кг/м3
Предлагаемое значение 607.84			
1.529	NaN	<input type="checkbox"/>	Стандартная сепарация - Объемный коэффициент нефти, м3/м3
Предлагаемое значение 2.15			
0.55	NaN	<input type="checkbox"/>	Вязкость пластовой нефти, мПа·с
Предлагаемое значение 0.37			
Кластер 0			
Аномальная проба: плотность 1.29e-06 (критич. знач. 5.04e-06)			

Рис. 2. Интерфейс инструмента проверки проб на аномальность. Пользователю предоставляется возможность зафиксировать те параметры, значениям которых пользователь доверяет. Если проба признается (по рекомендации инструмента) аномальной, пользователю предлагаются значения, наиболее правдоподобные в рамках определенного кластера (найденного ранее)

Экспертная система для оценки геофизического исследования скважин

Принятие ключевых решений о разработке месторождения напрямую зависит от интерпретации данных геофизических скважинных исследований. Интерпретация производится подразделениями компании-оператора месторождения, внешними подрядчиками и напрямую зависит от опыта экспертов (геолога, петрофизика) и знаний особенностей конкретного региона, лицензионного участка, объекта разработки. Сами геофизические данные содержат много неопределенностей, что усложняет задачу интерпретации и придает ей зачастую вероятностный характер. Для снижения неопределенностей необходимо использовать дополнительные независимые экспертизы. Таким независимым экспертом может быть система, построенная на алгоритмах машинного обучения. Если интерпретация машины будет отличаться от интерпретации эксперта, при последующем проектировании разработки необходимо провести дополнительные исследования или дополнительную экспертизу существующих данных.

Анализ данных геофизических исследований скважин (ГИС) уже осуществлялся с применением методов машинного обучения. Искусственные нейронные сети применялись для определения литофаций [2], частоты естественной трещиноватости [3] и пористости [4] вдоль ствола скважины. В [5] предложен метод идентификации коллекторов вертикальных скважин с учетом пространственного положения относительно других скважин на основе fPCA (Functional PCA, функциональный метод выделения главных компонент) и множественной линейной регрессии. Методы ближайшего соседа и опорных векторов применялись для интерпретации ГИС в случае кристаллических пород [6]. Рекуррентные нейронные сети применялись для выделения коллектора вдоль ствола скважины [7].

Ниже рассмотрен подход с точки зрения задачи сегментации, что позволяет значительно снизить влияние шума в исходных данных и одновременно учесть большие интервалы глубин. Среди исследованных алгоритмов, показывающих максимально качественные и устойчивые результаты сегментации, можно выделить одномерную сверточную нейронную сеть. Архитектура использованной одномерной сверточной нейронной сети представляет собой схему encoder-decoder.

Архитектура построена на основе сети U-Net [8], где одна половина создана для извлечения полезных признаков из данных (encoder, «кодирующая»), а вторая — для построения на их основе результата сегментации (decoder, «восстанавливающая»). Особенностью архитектуры является «проброс» слоев между «кодирующими» и «восстанавливающими» слоями, что позволяет точнее воспроизводить границы сегментированных объектов.

В качестве экспериментальных данных использованы каротажные сигналы из более 200 вертикальных

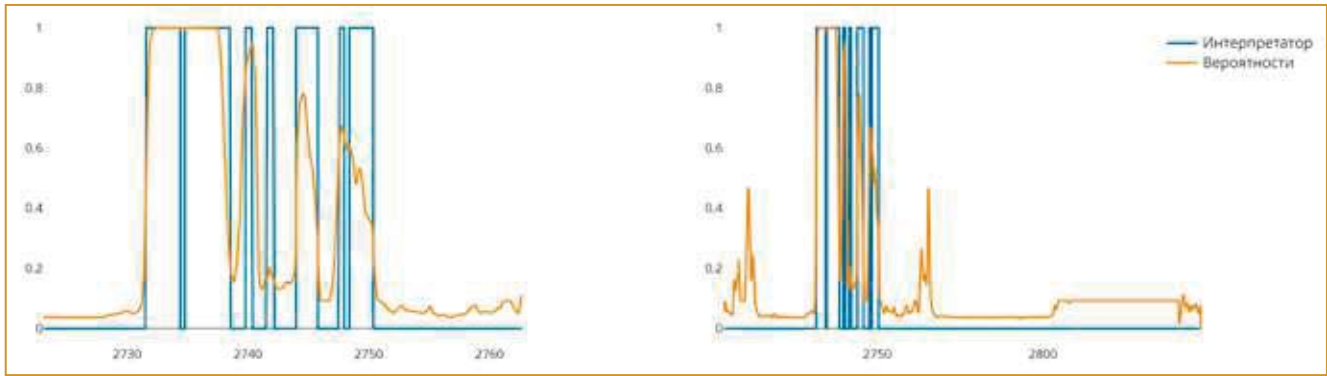


Рис. 3. Слева - зависимость вероятности (априорной) наличия коллектора (зеленая кривая), полученные на основе сверточной нейронной сети, и исходная интерпретация (синяя кривая); справа - зависимость, построенная для всего исследованного интервала глубин

скважин, относящихся к одному месторождению. Принципиальными для обучения моделей являются следующие особенности исходных данных: классы («коллектор», «не коллектор») не сбалансированы; среди каротажей только пять проведены почти во всех скважинах; большая часть каротажей содержит выбросы

После предобработки данных из выборки проделаны стандартные процедуры: скважины случайным образом разделены на основную, валидационную и тестовую выборки. Разбиение проходило из расчета (приблизительно) 80% на основную, 20% на валидационную и 1% на тестовую (1 скважина). На (рис. 3) представлен пример предсказания вероятности наличия коллектора по данным ГИС с использованием сверточной нейронной сети.

Качество предсказания по метрике F1 составляет 0,71, что не позволяет полностью автоматизировать процесс, но дает альтернативную точку зрения на вариант разметки эксперта, обращая внимание на моменты, содержащие неопределенность или неточность. Пример показан на рис. 4 (справа).

Автоматизация петрографического анализа цифровых изображений шлифов терригенных пород

Геофизические исследования скважин не являются самодостаточными. Разрешающая способность большинства типов исследований составляет десятки сантиметров, что приводит к усреднению профилей характеристик вдоль скважины. При этом результаты измерения конкретного физического поля могут существенно зависеть от производителя оборудования, условий измерения, способа предобработки и т. д. Для преодоления данных ограничений каротажные данные дополняются исследованиями кернавого материала, извлеченного из отдельных скважин.

Среди кернавых исследований анализ полноразмерного керна и петрографический анализ шлифов (небольших образцов) являются наименее сложными при реализации. Большая часть месторождений нефти и газа России приурочены к терригенным отложениям. На микро-масштабе они состоят из обломков (зерен), которые образуются при разрушении горных пород. Сегментация, классификация и последующий расчет свойств породообразующих зерен важен для диагностики и описания осадочных обломочных пород, на основе которых проводятся стратиграфические корреляции, историко-геологические реконструкции и комплекс геологоразведочных работ, а также характеристика резервуаров углеводородов. При петрографическом анализе алевропесчаных пород оцениваются такие гранулометрические характеристики, как линейный размер, параметры формы и упаковки породообразующих зерен, со-

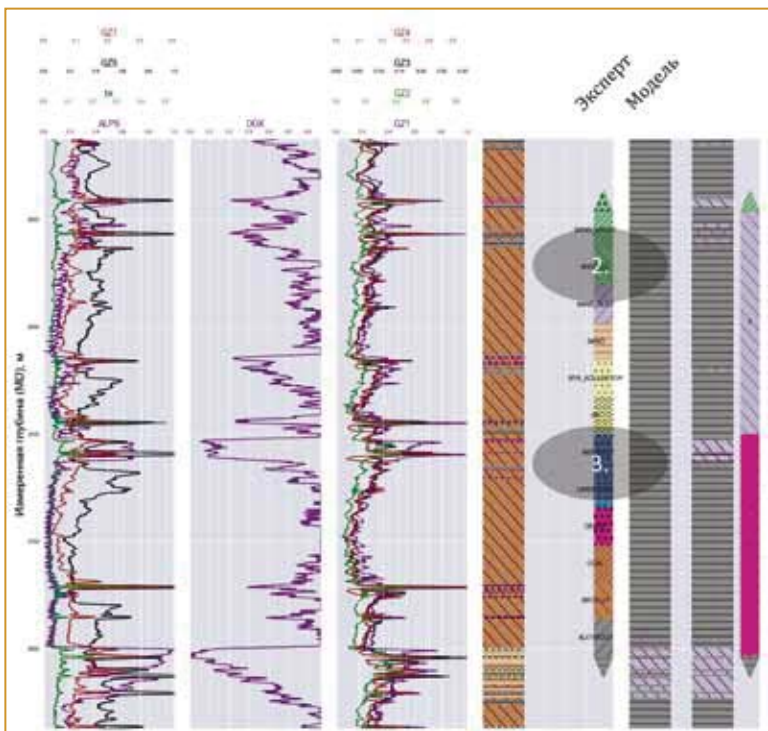


Рис. 4. Пример экспертной системы. Области 1 и 2 соответствуют несовпадениям предсказания модели и эксперта, соответственно, требуют дополнительной верификации

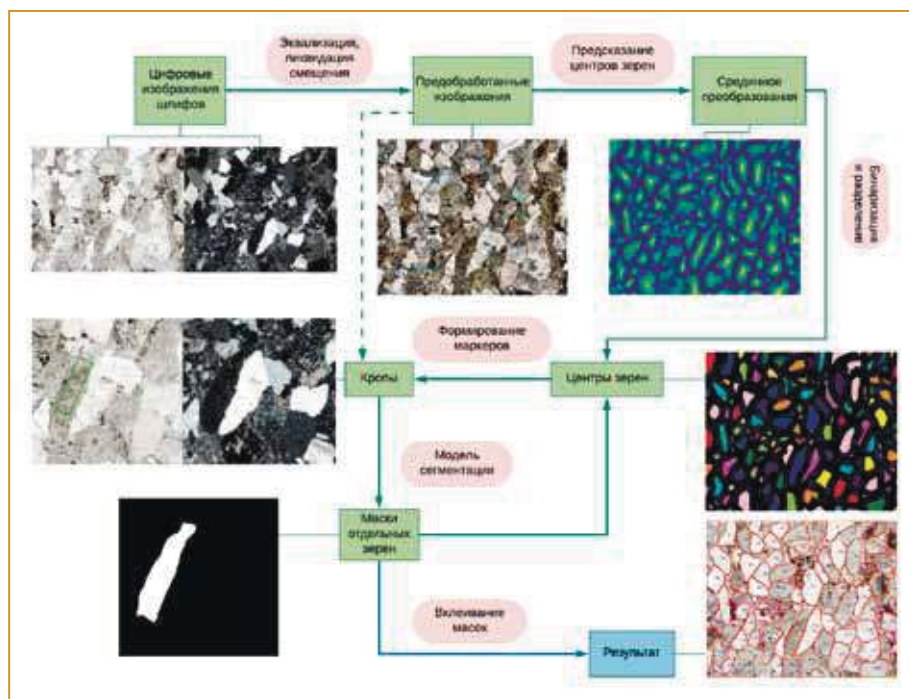


Рис. 5. Верхнеуровневая архитектура автоматической сегментации порообразующих зерен

ртировка, модальный и медианный диаметры. Расчет с помощью микрометра производится на основании 200...300 зерен, могут использоваться различные методические подходы. С целью унификации петрографических данных, повышения их качества, сокращения временных затрат на подсчет зерен была поставлена задача автоматизации рутинных процедур.

В целях повышения качества данных и ускорения петрографического анализа решена задача автоматической обработки цифровых изображений шлифов терригенных пород с применением гибридного подхода IP&DL (Image Processing & Deep Learning). Ключевой решенной задачей является полностью автоматическая сегментация всех структурных объектов (порообразующие зерна, пустотное про-

странство, цемент) [9]. Реализованный алгоритм позволяет провести количественный анализ структуры породы [10]. В частности, предоставляется возможность по цифровым изображениям шлифов терригенных пород провести количественную оценку размерности обломочных зерен (гранулометрический состав), расчет морфологических свойств отдельных зерен (окатанность, сферичность/удлиненность), относительной площади структурных объектов, процентное соотношение площадей различных классов зерен (минералов); описать текстурные свойства образца (ориентацию зерен, тип контактов зерен, анизотропию и однородность расположения порообразующих зерен).

Для решения задачи сегментации и классификации структурных объектов профильными экспертами подготовлена выборка из более чем 9000 отдельных порообразующих зерен. Для каждого зерна указан предполагаемый минеральный состав, тип вторичных изменений, аппроксимированный полигоном контур. Собранная база данных позволила апробировать существующие подходы решения задачи сегментации порообразующих зерен, реализовать и протестировать свой. Важно отметить, что реализация собственного алгоритма сегментации зерен потребовалась по причине отсутствия задач-аналогов. Существенная часть зерен сильно вторично изменены, что вносит степень неопределенности при локализации объектов и их границ. При этом объекты чаще всего плотно упакованы, что затрудняет использование существующих алгоритмов детекции [11] для локализации.

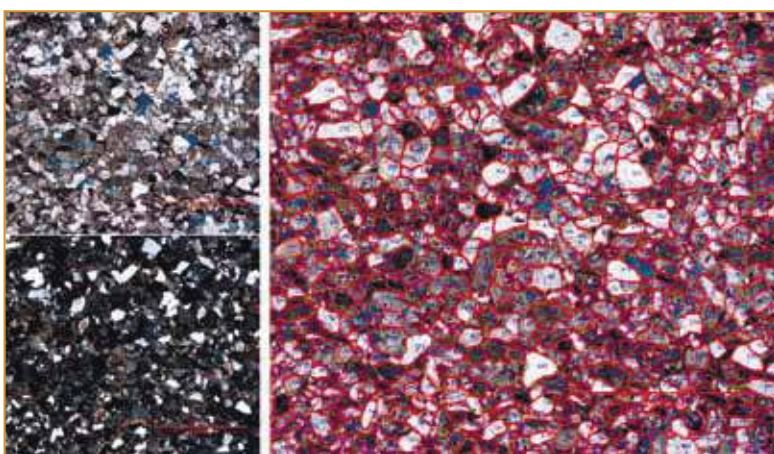


Рис. 6. Пример корректной сегментации в условиях сильных вторичных изменений пород при относительно небольших размерах зерен (средний диаметр Ферета < 32 пикселей)

Верхнеуровневая схема решения представлена на рис. 5. Алгоритм построен на двух моделях сверточных нейронных сетей: “локализирующей” и “уточняющей”. На основании предобработанных входных изображений первая модель восстанавливает нормализованное медиальное преобразование. В предположении, что объекты квази-выпуклы, посредством бинаризации из полученной поверхности могут быть извлечены центры зерен. Отдельные центры используются в качестве маркеров для второй модели, которая внутри цикла восстанавливает бинаризованную маску для каждого зерна отдельно. Подробное описание подхода и результатов тестирования приведены в [12].

Тестирование подхода проводилось как на выборке с разметкой для количественной оценки результатов, так и на расширенной выборке без разметки для оценки наименее корректных предсказаний — качественной оценки. Обнаруженные ошибки далее исправлялись, а сам образец добавлялся в обучающую выборку. Оценка качества на размеченной выборке проведена посредством кросс-валидации ($k=10$). Средние значения метрик качества AP и F1 составили 0,76 и 0,89 соответственно. Качественная оценка показала, что для одного образца необходимо исправлять (корректировать границы, удалять, добавлять) 5...15 зерен (на 300 корректных). Подобное качество позволяет существенно (в 20...60 раз) ускорить процесс сегментации пороодообразующих зерен в образцах специалистом и делает возможным автоматический расчет морфологических характеристик и свойств упаковки зерен. Пример изображения образца из расширенной выборки с автоматически обведенными границами приведен на рис. 6.

Выводы

Наращивание объема хранимых данных нефтегазового промысла создает потенциал для внедрения современных методов анализа данных, таких как методы машинного/глубинного обучения. Данные технологии, основанные на динамично развивающемся направлении анализа больших данных, позволяют повысить рентабельность добычи. Редко данное увеличение будет кратным, но те проценты ошибок, которые алгоритмы машинного обучения помогут избежать человеку, принимающему ключевые решения, могут оказаться критичными для успешности проектов. Время, которое сэкономят высококвалифицированный специалист, переложив на машину наименее творческую часть работы, может быть использовано для более релевантной задачи. При этом обучающаяся машина может дать дополнительное мнение стороннего эксперта при интерпретации сложных данных. Нарастающее присутствие технологий машинного обучения в обработке информации и их применение совместно с традиционными численными моделями представляется несомненным трендом и необходимостью сегодняшнего дня. В настоящей работе продемонстрирован пример успешного использования методов машинного/глубинного обучения для решения критически важных задач в разрезе анализа геолого-

геофизических данных нефтяного промысла: выявление некорректных показателей в базе данных свойства пластовых флюидов, определение неоднозначных с точки зрения интерпретации ГИС интервалов в рамках задачи поиска коллектора, автоматизация анализа морфологических признаков структурных объектов терригенных пород по цифровым снимкам изображений петрографических шлифов.

Список литературы

1. *Bishop C.M. Pattern Recognition and Machine Learning. 2006. 738.*
2. *Bhatt, Alpiana, and Hans B. Helle. Determination of facies from well logs using modular neural networks // Petroleum Geoscience 8.3 (2002): 217-228.*
3. *FitzGerald, Elaine M., Christopher J. Bean, and Ronan Reilly. Fracture-frequency prediction from borehole wireline logs using artificial neural networks // Geophysical prospecting. 47.6 (1999): 1031-1044.*
4. *Wong, Patrick M., Tamás D. Gedeon, and Ian J. Taggart. An improved technique in porosity prediction: a neural network approach // IEEE Transactions on Geoscience and Remote Sensing. 33.4 (1995): 971-980.*
5. *Kormaksson, Matthias, Marcos R. Vieira, and Bianca Zadrozny. A data driven method for sweet spot identification in shale plays using well log data // SPE Digital Energy Conference and Exhibition. Society of Petroleum Engineers. 2015.*
6. *Konaté, Ahmed Amara, et al. Machine Learning Interpretation of Conventional Well Logs in Crystalline Rocks. // International Conference in Swarm Intelligence. Springer. Cham. 2015.*
7. *Belozеров, Boris, et al. Automatic Well Log Analysis Across Priobskoe Field Using Machine Learning Methods. SPE Russian Petroleum Technology Conference. Society of Petroleum Engineers. 2018.*
8. *Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention. Springer. Cham. 2015. p. 234-241.*
9. *Bukharev, A. Y., et al. Automatic Analysis of Petrographic Thin Section Images Of Sandstone // ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery. 2018.*
10. *Bukharev A. Y. et al. Digitization of Petrographic Data for Standardization and Genetic Interpretation. SPE Russian Petroleum Technology Conference. Society of Petroleum Engineers. 2018.*
11. *Zhang, Shifeng, et al. Single-shot refinement neural network for object detection. arXiv preprint (2017).*
12. *Bukharev A. Y., et al. The task of instance segmentation of mineral grains in digital images of rock samples (thin sections) // International Conference on Artificial Intelligence Applications and Innovations (IC-AIAI). 2018.*

Буденный Семен Андреевич — руководитель департамента цифровых технологий в индустрии,

Волков Никита Алексеевич — аналитик данных,

Цанда Артём — младший инженер, **Бухарев Александр Юрьевич** — старший аналитик данных департамента цифровых технологий в индустрии ООО «Инжиниринговый центр МФТИ».

Контактный телефон +7 (498) 744-66-83. E-mail: Budennyu.sa@cet-mipt.ru

Маргарит Андрей Станиславович — начальник отдела научно-технического развития,

Андреанова Алла Михайловна — ведущий специалист управления комплексного проектирования разработки,

Жуковская Елена Анатольевна — ведущий эксперт,

Белозеров Борис Владимирович — начальник управления — ведущий эксперт управления геологической экспертизы и научно-методического развития ООО «Газпромнефть НТЦ».

Контактный телефон +7 (812) 313-69-24. E-mail: Belozerov.BV@gazpromneft-ntc.ru