

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ СОЗДАНИЯ СИСТЕМЫ ПРЕДОТВРАЩЕНИЯ ОСЛОЖНЕНИЙ ПРИ БУРЕНИИ НЕФТЕГАЗОВЫХ СКВАЖИН

УДК 004.032.26::622.248.54

С.А. Косяченко, к.т.н., ООО «Цифра» (Москва, РФ),

Semen.Kosyachenko@zyfra.com

А.Н. Тарасова, ООО «Цифра», Anna.Tarasova@zyfra.com

Возможность раннего прогнозирования прихватоопасных состояний в процессе строительства скважин – важная задача для нефтегазовой отрасли. В настоящее время, как правило, мониторинг осложнений проводят посредством сравнения фактических и расчетных показателей без использования искусственного интеллекта.

В представленной статье проанализированы перспективы методов машинного обучения для поиска и использования скрытых зависимостей между наблюдаемыми параметрами бурения в целях заблаговременного информирования специалистов о возможных прихватах. Рассмотрены вопросы выбора пространства (набор параметров) для оценки состояния скважины и реализации моделей машинного обучения. Приведены методы оценки общего состояния (аномальности) скважины на основе кластеризационной модели (выделение состояний скважин, похожих на предприхватные и прихватные, из истории наблюдений) и оценки дистанции по Махаланобису [1]. Описаны способы расчета ключевых показателей (вес на крюке, момент на роторе, давление бурового раствора на входе) с использованием регрессионных моделей сравнения («модель – факт»), реализованных посредством модели «случайный лес». Рассмотрена возможность построения прогностических моделей, базирующихся только на управляемых в процессе бурения параметрах, на основе линейных регрессионных моделей. Приведена общая схема функционирования сервиса, объединяющая отдельные модели машинного обучения. Результаты численных исследований с применением архивных данных геолого-технических исследований позволяют высоко оценить перспективы моделей машинного обучения применительно к проблемам бурения нефтегазовых скважин.

КЛЮЧЕВЫЕ СЛОВА: БУРЕНИЕ, ГЕОЛОГО-ТЕХНИЧЕСКОЕ ИССЛЕДОВАНИЕ, ДИСТАНЦИЯ ПО МАХАЛАНОБИСУ, КЛАСТЕРИЗАЦИОННАЯ МОДЕЛЬ, МАШИННОЕ ОБУЧЕНИЕ, МОДЕЛЬ SHAP VALUES, ОСЛОЖНЕНИЕ, ПРИХВАТ, РЕГРЕССИОННАЯ МОДЕЛЬ.

Процесс строительства скважин в последнее время претерпевает некоторые изменения. Вследствие постепенного истощения продуктивных пластов необходимо бурить горизонтальные скважины с большим отходом от вертикали, в некоторых случаях достигающим 10–12 тыс. м по стволу. В связи с более сложной проводкой возрастает риск осложнений и аварийных ситуаций. По этой причине в последнее время вопросу безопасного ведения работ уделяется много внимания.

Наиболее распространенные осложнения – прихваты, которые возникают по различным причи-

нам. Ниже перечислены основные из них:

- перепад давления;
- образование сальников;
- заклинивание бурильного инструмента в интервалах сужения ствола, при обвалах неустойчивых пород.

Осуществление буровых и спуско-подъемных операций всегда сопровождается информационной поддержкой. В то же время перед сервисными компаниями все еще стоит задача по созданию системы, способной заблаговременно предупреждать о вероятных осложнениях.

На нефтегазовых предприятиях ведутся работы по совершенствованию технологий в целях предотвращения прихватов: подбирается специальный состав промывочной жидкости для вскрытия пород с различными физико-механическими свойствами, выбираются траектории ствола с минимизацией извилистости, рассчитываются оптимальные параметры процесса, проводится мониторинг в реальном времени путем сравнения фактических и расчетных показателей для определения отклонений. Такие действия основываются на статистике и опыте конкретного специа-

S.A. Kosyachenko, PhD in Engineering, 000 Digit [000 Tsifra] (open joint-stock company) (Moscow, the Russian Federation), Semen.Kosyachenko@zyfra.com

A.N. Tarasova, 000 Digit, Anna.Tarasova@zyfra.com

Using machine learning techniques to create a complication preventing system for oil and gas well drilling

Early prediction of sticky states during well construction is an essential task for the oil and gas industry. To date, complication monitoring commonly uses comparing actual and design figures and does not exploit the potential of artificial intelligence. The article analyses the prospects of machine learning techniques for finding and using covert dependencies between the observed drilling parameters to notify the technicians in advance about a possible sticking. The paper also addresses selecting space (parameter set) for well state assessment and implementation of machine learning models. The methods are given for the general well state (well anomaly) assessment based on a clustering model (identification of well states similar to sticky and pre-sticky ones, from the recorded history) and Mahalanobis distance [1] assessment. The study also describes the calculation methods for the key indicators (hook weight, rotary torque, surface mud pressure) using regression comparison models (model-actual) implemented by random forest. We have considered building predictive models based on linear regression models that only use parameters controlled during drilling. A general operation diagram is given for the service, comprising the separate machine learning models.

The results of the numerical study that uses the archived geotechnical well testing data allow us to appreciate the prospects of machine learning models applied to the issues of the oil and gas well drilling.

KEYWORDS: COMPLICATION, DISTANCE, DRILLING, GEOTECHNICAL WELL TESTING, MACHINE LEARNING, MAHALANOBIS CLUSTERING MODEL, REGRESSION MODEL, SHAP VALUES MODEL, STICKING.

листа. К сожалению, предлагаемые решения для интеллектуальной поддержки строительства скважин развиваются не столь активно, хотя применение методов машинного обучения представляется крайне перспективным.

Среди крупных поставщиков услуг и программных продуктов в нефтегазовой отрасли можно отметить Roxar AS (Норвегия), которая с помощью пакета DrillScene способна детектировать тренды, позволяющие оперативно диагностировать признаки возможных осложнений [2]. Специалисты ПАО «Газпром нефть» развивают систему, предназначенную для прогнозирования аварийных ситуаций и своевременного информирования сотрудников об их возникновении [3]. В открытом доступе можно найти ряд научно-исследовательских работ по данной теме [4, 5 и др.].

Несмотря на существующие разработки, проблема безаварийного бурения скважин не может считаться решенной. Перспективы применения методов машинного обучения в данной области и необходимость создания конкурентоспособного российского программного обеспечения, пред-



назначенного для своевременного информирования специалистов о вероятных осложнениях, послужили толчком к созданию описываемой в данной статье системы.

ПОСТРОЕНИЕ ПРОСТРАНСТВА ДЛЯ РЕАЛИЗАЦИИ МОДЕЛЕЙ

Прихваты могут возникать по различным причинам, которые характеризуются отклонением от нормы определенного набора параметров. Для предприхватных состояний свойственно:

- увеличение/уменьшение веса на крюке при подъеме/спуске буровой колонны (вследствие сальникообразования, попадания постороннего предмета в сква-

жину, нарушения устойчивости ее стенок, желобообразования в зонах сужения ствола);

- рост момента на роторе (в зонах сужения ствола, из-за оседания породы на инструменте);

- большие скачки момента на роторе (частичное подклинивание инструмента);

- повышение давления на входе (сужение ствола, сальникообразование, нарушение устойчивости стенок скважины).

Построение системы прогнозирования предприхватных и прихватных состояний может быть основано на распознавании приведенных выше шаблонов. Для успешного обнаружения любого из них важно

правильно выбрать пространство (набор параметров), в котором этот искомый шаблон будет наилучшим образом отличим от остальных состояний. Основой для такого пространства стали измерения параметров геолого-технологических исследований (ГТИ):

- положение талевого блока;
- глубина долота;
- глубина забоя;
- момент на роторе;
- обороты ротора;
- давление на входе;
- вес на крюке;
- механическая скорость.

Кроме приведенных выше, были использованы дополнительные параметры (ДП), которые представляют собой функции исходных параметров и истории их измерений. Список функций для построения ДП приведен ниже.

Скользящее среднее по параметру M на выбранном временном периоде:

$$M(P, t, \Delta t) = \frac{1}{n} \sum_{t_i = t - \Delta t}^t P(t_i), \quad (1)$$

где P – исходный параметр; t – момент времени, для которого восстанавливается значение ДП; Δt – период усреднения; n – количество замеров в период $[t_i, t]$; $t_i = t - \Delta t$.

Изменение параметра C за выбранный временной период:

$$C(P, t, \Delta t) = \frac{P(t) - P(t_i)}{\Delta t}. \quad (2)$$

Значение параметра L с временным сдвигом:

$$L(P, t, \Delta t) = P(t_i). \quad (3)$$

Также ДП были реализованы с помощью различных физических моделей. В частности, большой вклад в качество итогового решения внесла упрощенная физическая модель коэффициента трения в скважине [6]:

$$F(t) = F_0 \frac{R(t)}{\sqrt{R(t)^2 + O(t)^2}}, \quad (4)$$

где $F(t)$ – расчетное значение коэффициента трения на долоте

Таблица 1. Основные и дополнительные параметры, использованные в рассмотренных моделях машинного обучения
Table 1. Principal and secondary parameters used in the machine learning models in question

Исходный параметр P Source parameter P	Порождающая функция и ее параметры Generating function and its parameters	Результирующий параметр Resulting parameter
Давление на входе d Surface pressure d	—*	$d(t)$
Момент на роторе m Rotary torque m	—	$m(t)$
Вес на крюке h Hook weight h		$h(t)$
Давление на входе d Surface pressure d		$M(d, t, \Delta t)$
Момент на роторе m Rotary torque m	$M, \Delta t = 60 \text{ c (s)}$	$M(m, t, \Delta t)$
Вес на крюке h Hook weight h		$M(h, t, \Delta t)$
Давление на входе d Surface pressure d		$C(d, t, \Delta t)$
Момент на роторе m Rotary torque m	$C, \Delta t = 10 \text{ c (s)}$	$C(m, t, \Delta t)$
Вес на крюке h Hook weight h		$C(h, t, \Delta t)$
Глубина долота g Bit depth g	$M, \Delta t = 20 \text{ c (s)}$	$M(g, t, \Delta t)$
Глубина забоя z Bottom-hole depth z		$M(z, t, \Delta t)$
Механическая скорость R , обороты ротора O Rate of penetration R , rotor speed O	F	$F(t)$

* Примечание. Порождающая функция может не применяться, в этом случае берется исходный параметр

* Note. Generating function may not be applied: in this case, the source parameter is used

в момент времени t ; F_0 – коэффициент трения покоя; $R(t)$ – скорость движения ротора; $O(t)$ – обороты ротора.

Реализация большинства физических моделей – отдельная сложная проблема, которая требует особого внимания и в рамках статьи не рассматривается.

Таким образом, применение к исходному параметру одной из приведенных выше функций позволяет построить большое число ДП, каждый из которых можно использовать для обогащения пространства измерения скважины с помощью порождающих (дополнительных) функций.

При выборе пространства состояний для реализации той или иной статистической модели необходимо помнить о том, что цель его построения – максимально всестороннее описание состояния скважины. Под последним здесь и далее будем понимать значения измеренных параметров ГТИ и величины ДП в момент времени t . Следует уточнить, что употребление множества ДП, построенных на одном исходном параметре, позволяет получить большой объем информации, однако приводит к сосредоточению внимания моделей на самих ДП. Для предотвращения этой про-

блемы необходимо тщательно с аналитической точки зрения подходить к выбору параметров для построения пространства состояния.

Для каждой модели пространство состояния строили индивидуально, основываясь на аналитических данных, здравом смысле и результатах численных экспериментов. Тем не менее можно утверждать, что ядро пространства моделей – это множество параметров, приведенных в табл. 1.

ТИПЫ МОДЕЛЕЙ

В ходе работы над данной проблемой исследованы возможности решения (или частичного решения) задачи прогнозирования осложнений. В результате выработано три основных подхода:

- модель оценки общей аномальности состояния скважины;
- модель сравнения («модель – факт»);
- классическая прогнозная модель.

Рассмотрим приведенные типы моделей подробнее.

МОДЕЛЬ ОЦЕНКИ ОБЩЕЙ АНОМАЛЬНОСТИ СОСТОЯНИЯ СКВАЖИНЫ

Модели данного типа основаны на предположении о том, что скважины большую часть времени находятся в нормальном состоянии. Под нормальным будем понимать состояние, исходные параметры которого отличаются от проектных не более чем на заданное допустимое отклонение. В общем случае можно считать, что в процессе бурения скважина почти всегда находится в нормальном состоянии. За аномальные примем состояния, наблюдающиеся впервые или очень редко. Для них характерны последующие восстановительные/аварийные работы.

Авторами исследованы две модели детектирования аномальности состояния скважины:

- кластеризационная;
- основанная на вычислении дистанции по Махаланобису.

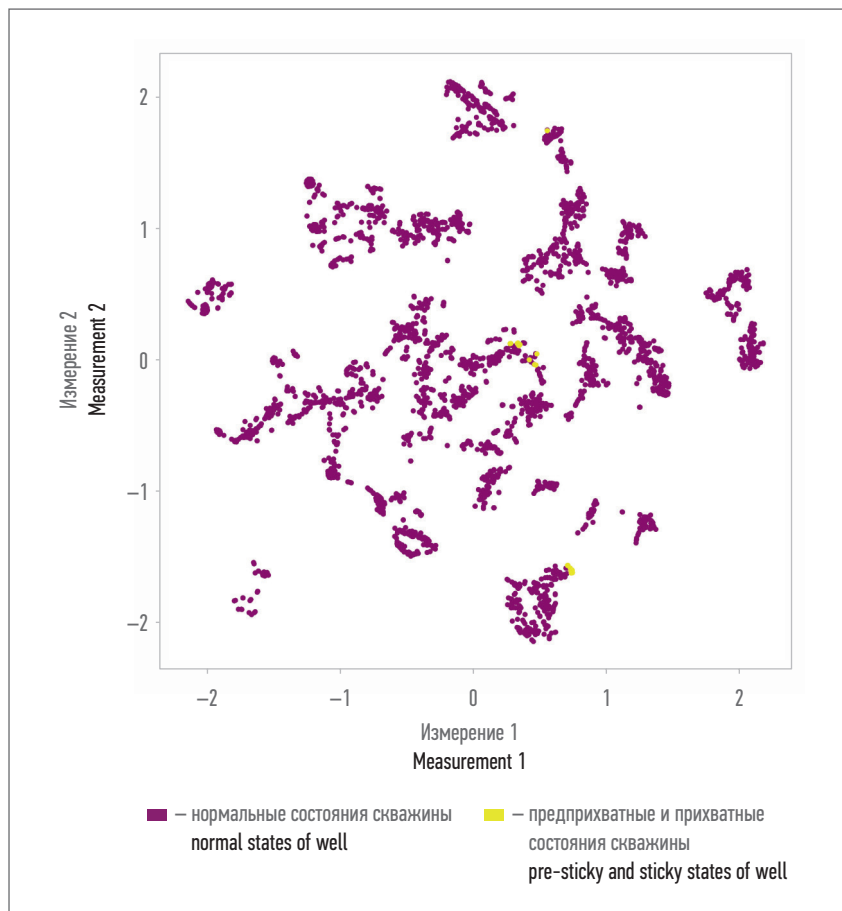


Рис. 1. Двухмерная проекция t-SNE-кластеризации состояний скважины
Fig. 1. Two-dimensional projection of t-SNE clustering of well states

Кластеризационная модель

В качестве кластеризационной использовали модель t-SNE (t-distributed stochastic neighbor embedding) [7], поскольку она дает наилучшую различимость между нормальными и аномальными состояниями скважины.

Исследовали девять скважин одного из месторождений в Западной Сибири. Историю бурения каждого объекта разделили по операциям, далее детально рассмотрели процесс бурения ротором. На одной из скважин обнаружили серьезное осложнение – механический прихват, от которого удалось освободиться только через несколько дней, с проведением восстановительных/аварийных работ. Предприхватные состояния этой скважины промаркировали (рис. 1). Далее обратили внимание на кластеры, в которые попали предприхватные и прихватные

состояния. После анализа выяснилось, что состояния, попавшие в предприхватный кластер, имеют признаки предприхватного состояния.

Несмотря на хороший потенциал, данный подход имеет существенный недостаток – высокие требования к качеству и объему разметки состояний скважины. Под разметкой понимается список, в котором каждому состоянию соответствует метка «прихват» или «не прихват». Однако реализация кластеризационной модели затруднительна, поскольку на новых месторождениях необходимо актуализировать имеющуюся разметку.

Дистанция по Махаланобису

Дистанция по Махаланобису – мера расстояния между случайными величинами. Она отличается от Евклидова расстояния тем,

Таблица 2. Оценка полноты моделей
Table 2. Model recall assessment

Модель Model	Полнота модели Q_p , % Model recall Q_p , %			
	За 12 ч до прихвата 12 h before sticking	За 4 ч до прихвата 4 h before sticking	За 30 мин до прихвата 30 min before sticking	За 30 мин до затяжки/посадки 30 min before drag/slack-off
Общая аномальность по Махаланобису General anomaly based on Mahalanobis distance	80,0	100,0		56,4
PM1. Вес на крюке PM1. Hook weight	–	60,0	80,0	65,2
PM1. Момент на роторе PM1. Rotary torque		40,0	80,0	46,1

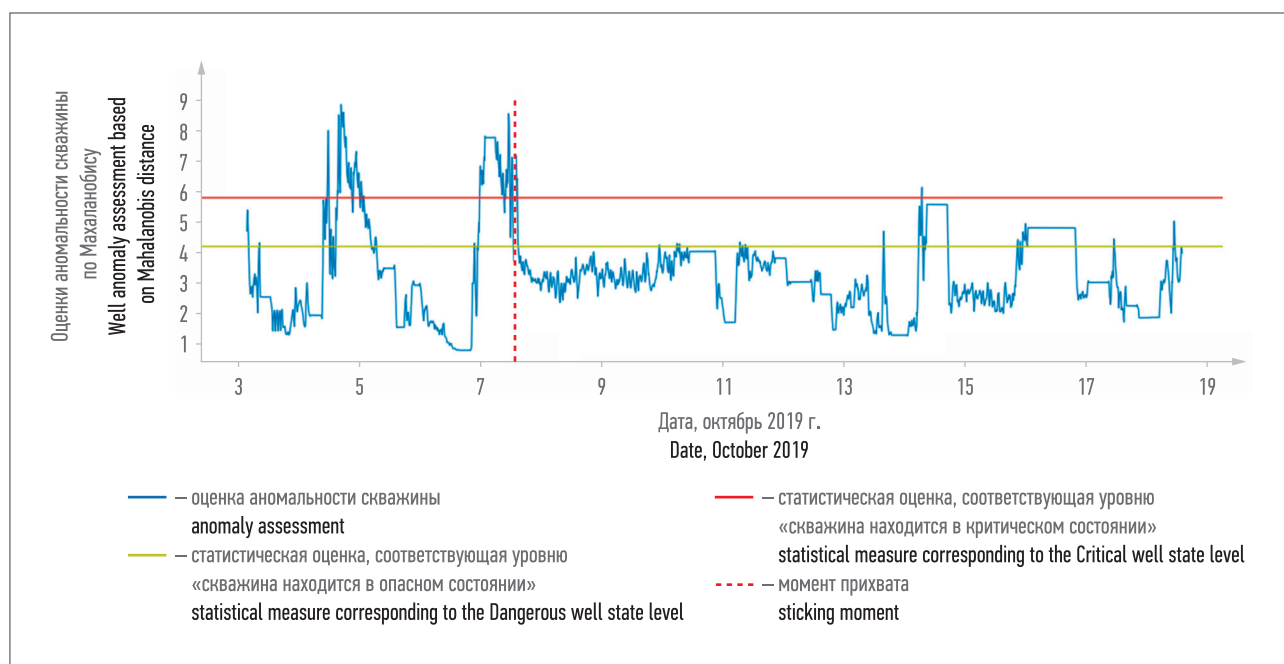


Рис. 2. Оценка аномальности скважины по Махаланобису во времени
Fig. 2. Well anomaly assessment based on Mahalanobis distance over time

что учитывает корреляции между измерениями и инвариантна к масштабу [1].

Для реализации оценки состояния по Махаланобису нужно накопить историю наблюдений за скважиной для построения ковариационной матрицы. Таким образом, ограничение данной модели состоит в необходимости сбора информации в течение некоторого времени. Для типичных условий этот период составляет 8–30 ч. В случаях, когда начинается бурение секции и данные еще не получены, допустимо использовать архивную информацию о бурении похожих

участков с тем же оборудованием, однако это ведет к некоторому усложнению предлагаемой модели.

Реализованная модель расстояния по Махаланобису имеет высокую полноту оценки предприхватных состояний скважины (табл. 2). Согласно классическому определению полнота – это отношение количества найденных моделью аномальных состояний к числу всех аномальных состояний в истории; точность – отношение количества аномальных состояний в истории к числу состояний из этого подмножества, которые модель отнесла к аномальным [8].

На рис. 2 проиллюстрирована работа данной модели на примере одной из скважин. Согласно отчету, полученному от представителя подрядчика, осуществляющего наклонно направленное бурение, в некоторый момент времени T произошел прихват, от которого удалось освободиться только после 30 мин аварийных работ.

К недостаткам моделей данного типа можно отнести относительно невысокую точность: определяются моменты времени, когда модель срабатывает (показывает высокий уровень аномальности), т.е. скважина согласно расчетам находится

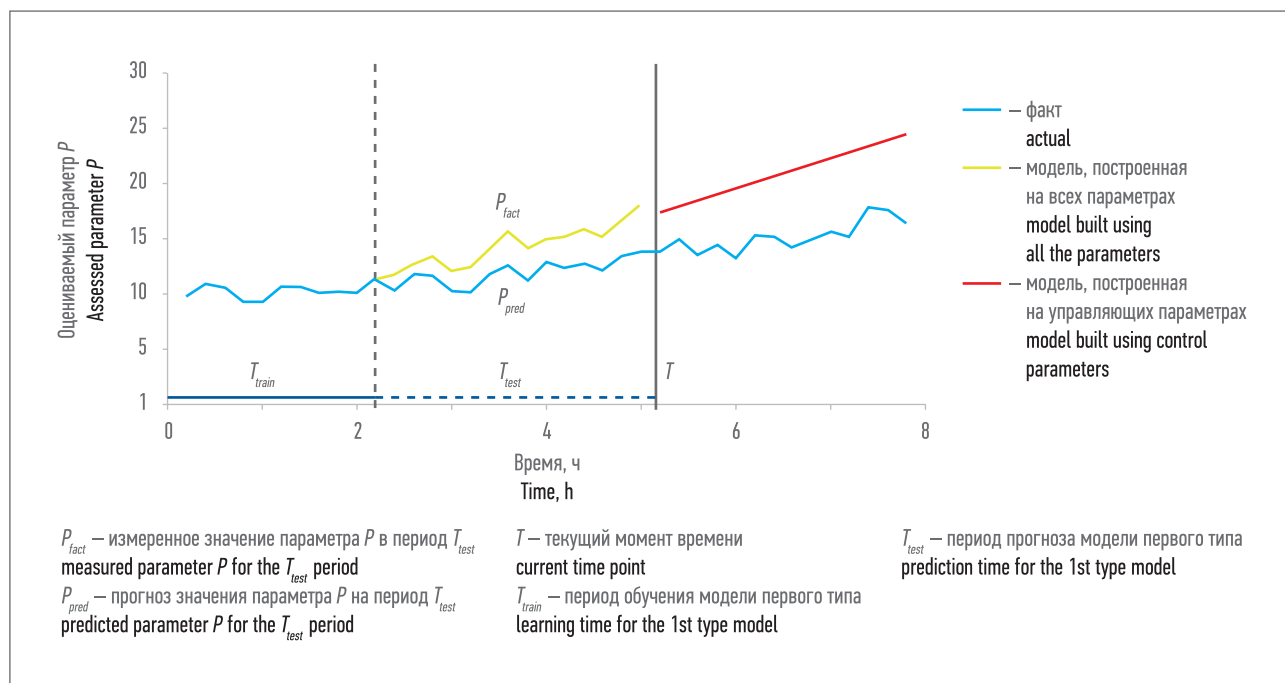


Рис. 3. Качественный график работы регрессионных моделей двух типов
Fig. 3. Quality plot for the two types of regression models

в опасном или критическом состоянии, хотя фактически осложнений при этом не наблюдается. В то же время не всегда можно точно сказать, находилась ли скважина действительно в опасном состоянии или нет. В любом случае для прогнозирования осложнений будем не только опираться на результат работы данной модели, но и подкреплять его другими моделями оценки состояния скважины или отдельных параметров.

РЕГРЕССИОННЫЕ МОДЕЛИ

Для оценивания параметров скважины в данной работе реализованы регрессионные модели двух типов:

- регрессионная модель сравнения «модель – факт» (PM1);
- прогнозная регрессионная модель (PM2).

Разберем эти два типа (рис. 3). Оценим значение параметра P в текущий момент времени. Тогда регрессионная модель первого типа – это модель, построенная на множестве наблюдаемых параметров за период наблюдения T_{train} . Ее прогнозом будет значение

параметра P (P_{pred}) за время T_{test} . Если предположить, что бурение в период T_{train} прошло без осложнений, то обученная на нем модель должна довольно точно предсказать значение P_{pred} . Если же P_{pred} существенно расходитсся с фактической величиной, можно сказать, что бурение в период T_{test} качественно отличается от предыдущего процесса. Другими словами, физические процессы в период T_{test} не совпадают с теми, которые наблюдались за период T_{train} (например, появилась нестабильность пробуренного ствола, что вызвало дополнительное трение).

В целях определения лучшей модели для решения поставленной задачи наиболее перспективным представляется использование рекуррентных искусственных нейронных сетей с долгой краткосрочной памятью (long short-term memory recurrent neural networks, LSTM) [9]. При применении LSTM-сетей в качестве модели прогнозирования наблюдались трудности настройки. Искомое состояние – прихват – не наступает мгновенно, к нему ведет цепь событий, которые растянуты во времени.

Поскольку LSTM хранит в том числе короткую память, она адаптируется к аномальным состояниям, и точность детектирования момента прихвата существенно снижается. В данном случае лучше работает модель «случайный лес» (random forest regression) [10], основанная на идее использования большого ансамбля решающих деревьев [11], каждое из которых дает невысокое качество классификации, но за счет их количества результат получается хорошим. Также возможно проведение бустинга [12] ансамбля деревьев, хотя это и не существенно улучшает адекватность модели.

Модели PM1 были реализованы для оценки трех параметров: момента на роторе, веса на крюке, давления на входе. Они характеризуются высокой полнотой и точностью (см. результаты численных исследований). В данной работе это основной класс моделей. В то же время следует отметить недостаток PM1: невозможность построения прогноза, поскольку значения наблюдаемых параметров после момента времени T неизвестны.

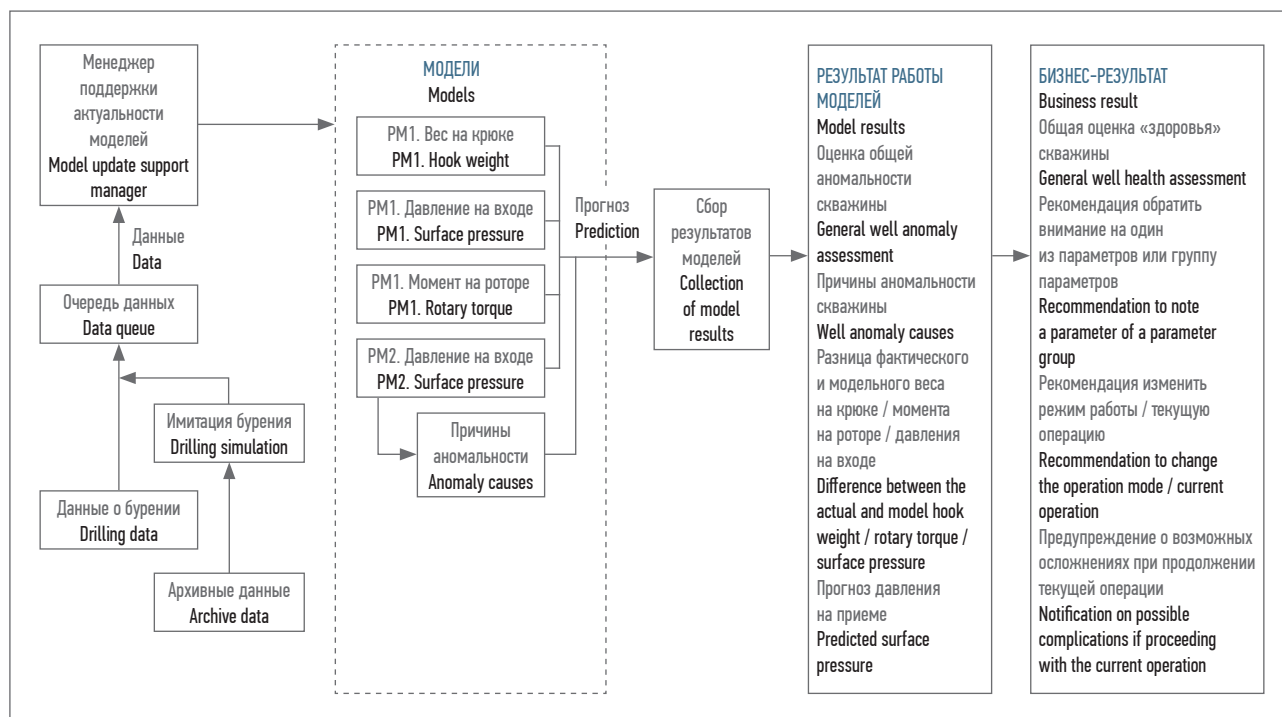


Рис. 4. Схема функционирования сервиса оценки состояния и построения рекомендаций
Fig. 4. Operation diagram for the state assessment and recommendation service

Для прогнозирования далее момента времени T предлагается использовать регрессионные модели второго типа (PM2). Они позволяют строить прогноз за момент времени T , однако в текущий момент наблюдений остается неизвестным, как будет происходить управление бурением или спуско-подъемными операциями и значения других наблюдаемых параметров. Таким образом, в текущий момент времени невозможно построить прогноз, не задавшись некоторыми предположениями. Эти предположения могут относиться только к управляющим параметрам бурения и не касаются остальных (наблюдаемых) параметров. Задача прогнозирования формулируется следующим образом. *В настоящий момент происходит операция X_0 с нагрузкой на долото X_1 , оборотами X_2 и расходом промывочной жидкости X_3 . Предположим, что текущая операция (X_0) продолжится в таком же режиме в течение времени X_4 . Каково будет давление на входе по окончании периода X_4 ? Такая постановка имеет ряд недостатков:*

- нет возможности использовать параметры ГТИ, за исключением управляющих, поскольку они неизвестны;
- подход применим только при относительной стабильности управляющих параметров в период $[t, t]$, т.е. тот период времени, в течение которого нужно определить текущий режим; если при этом происходит смена операции или меняется скорость/нагрузка/интенсивность, невозможно понять, как далее будет реализовано управление;
- низкая точность ввиду использования только управляющих параметров.

Вследствие ограниченности входных данных для построения модели (только предположение о значении управляющих параметров) для реализации PM2 можно использовать только статистически слабые модели. В данной работе применялась линейная регрессия [13]. Это позволило качественно понять тренд изменения интересующего параметра и избежать переобучения.

АСЕМБЛИРОВАНИЕ МОДЕЛЕЙ

Общая схема функционирования моделей представлена на рис. 4. Измерения в систему поступают из очереди, которая пополняется данными текущего бурения или сервисом имитации бурения. Данные собираются в менеджере поддержки актуальности моделей и с некоторой частотой подаются сервисам моделей для переобучения и получения прогноза/оценки. Далее все полученные результаты обрабатываются в сервисе сбора результатов с помощью специального алгоритма (не описан в рамках данной статьи).

Аномальность скважины отображается без обработки, при высокой аномальности производится анализ SHAP value (shapley additive explanations) [14] и индицируется ее причина.

Результаты работы моделей PM2 учитываются только в том случае, если полученный тренд превышает определенный заранее предел за заданное время; тогда формируется рекомендация изменить текущий режим для предотвра-

Таблица 3. Оценка точности моделей
Table 3. Model precision assesment

Модель Model	Точности модели Q_r , % Model precision Q_r , %			
	За 12 ч до прихвата 12 h before sticking	За 4 ч до прихвата 4 h before sticking	За 30 мин до прихвата 30 min before sticking	За 30 мин до затяжки/посадки 30 min before drag/slack-off
Общая аномальность по Махаланобису General anomaly based on Mahalanobis distance	34,4	47,0	67,3	23,1
PM1. Вес на крюке PM1. Hook weight	–	36,0	46,0	85,2
PM1. Момент на роторе PM1. Rotary torque		33,9	51,7	55,5

щения этого события. Примеры рекомендаций приведены ниже.

«Через 20 мин вероятны подклиники. Наблюдается тенденция на пилообразное изменение крутящего момента на роторе. Рекомендация: повышенный контроль за крутящим моментом».

«Через 30 мин вероятность прихвата составит 20 %. Наблюдается расхождение модельного и фактического веса на крюке при подъеме на 52 %; расхождение модельного и фактического давления на входе на 32 %. Рекомендация: повышенный контроль давления, крутящего момента, веса на крюке, расхода на входе и выходе».

ЧИСЛЕННЫЕ ИССЛЕДОВАНИЯ

Численные исследования представленных выше подходов осуществили, применив данные измерений, проведенных на 190 скважинах двух месторождений двумя разными заказчиками. Продуктивные пласты на обоих месторождениях залегают на глубине 4,0–5,5 тыс. м. Для определения точности работы моделей провели эксперименты с использованием исторических данных. В каждом из них моделировали бурение скважины – поступление данных ГТИ аналогично тому, как они поступали в процессе реального бурения. С момента старта имитации и далее через заданные промежутки времени происходило переобучение (актуализация) моделей – раз

в 15 мин, применение моделей – раз в 30 с. Результаты численных экспериментов по оценке полноты и точности основных моделей приведены в табл. 2, 3.

Измерения полноты и точности моделей общей аномальности проводили следующим образом. Выбрали 18 скважин, из истории измерения которых взяли 396 временных отрезков. Пять из них заканчивались прихватом, 56 – затяжкой или посадкой. На каждом отрезке модель обучалась на периоде $T_{train} = [T - 27 \text{ ч}, T - 13 \text{ ч}]$ и применялась на периоде $T_{test} = [T - 13 \text{ ч}, T - 12 \text{ ч}]$. Если результат модели более получаса находился выше заданной границы аномальности (определяется на основе анализа данных, полученных со множества скважин), считали, что скважина в этот час находится в аномальном состоянии; в противном случае состояние принимали нормальным.

Полноту Q_p в соответствии с [7] вычислили на основе тех экспериментов, которые оканчивались прихватом:

$$Q_p = \frac{1}{S} \sum_{i \in S} M_i \quad (5)$$

где S – подмножество экспериментов, которые заканчиваются прихватом; M_i – аномальность скважины для i -ого эксперимента.

$$M_i = \begin{cases} 1, & \left(\frac{1}{N} \sum_{t=t_{12}}^{t_{13}} m(t) \right) \geq M_i \\ 0, & \left(\frac{1}{N} \sum_{t=t_{12}}^{t_{13}} m(t) \right) < M_i \end{cases}, \quad (6)$$

где N – количество предсказаний на интервале $[t_{12}, t_{13}]$; t_{12}, t_{13} – границы исследуемого временного интервала (12, 13 ч до прихвата соответственно); $m(t)$ – полученное в результате работы модели значение аномальности в момент времени t ; M_i – определенная граница аномальности.

Результат описанного выше эксперимента можно видеть во второй колонке табл. 2. Точность моделей Q_r рассчитали по соотношению:

$$Q_r = \frac{\sum_{i \in S} M_i}{\sum_{i=1}^{396} M_i} \quad (7)$$

Оценки точности и полноты модели «PM1. Вес на крюке» получили аналогичным образом, с отличием в интерпретации результата работы модели M_i . Считали, что, если результат модели расходится с фактом более чем на 40 %, она способна предсказать прихват за 12 ч. Тогда M_i для данной модели составляет:

$$M_i = \begin{cases} 1, & \frac{1}{N} \sum_{t=t_{12}}^{t_{13}} \frac{|\hat{h}_j(t) - h_j(t)|}{h_j(t)} \geq 0,4 \\ 0, & \frac{1}{N} \sum_{t=t_{12}}^{t_{13}} \frac{|\hat{h}_j(t) - h_j(t)|}{h_j(t)} < 0,4 \end{cases}, \quad (8)$$

где $\hat{h}_j(t)$ – прогнозное значение веса на крюке в момент j ; $h_j(t)$ – фактический вес на крюке.

Аналогично провели эксперименты для остальных моделей и отрезков времени из табл. 2, 3.

Следует отметить, что каждая модель имеет свою зону

применимости. Например, «РМ2. Момент на роторе» нельзя использовать, если не происходит вращение инструмента, тогда как «РМ1. Вес на крюке» не функционирует при бурении. Для корректного анализа табл. 2 нужно иметь в виду, что основные проблемы ошибок моделей связаны с низким качеством данных или рассогласованиями.

ВЫВОДЫ

Проведенные численные исследования подтверждают применимость предлагаемого подхода

к оценке вероятности возникновения прихватов в процессе бурения скважин. Результаты, полученные в ходе данной работы, можно существенно улучшить путем использования данных других скважин. Также возможно повышение точности итогового прогноза и рекомендаций посредством анализа дополнительных моделей, в которых будут учитываться расход на входе и выходе, состав и характеристики породы, тип и реология бурового раствора. Следует отметить высокий потенциал кластеризационных моделей для долгосрочного про-

гноза – моделей, которые не будут в прямом смысле предсказывать изменение показателей, а будут фиксировать нестабильные участки скважины – потенциальный источник осложнений.

Положительные результаты исследований, представленные в статье, позволяют продолжить развитие моделей прогнозирования прихватов и реализацию программных продуктов, которые станут незаменимыми помощниками специалистам, непосредственно занимающимся бурением нефтегазовых скважин. ■

ЛИТЕРАТУРА

1. Prabhakaran S. Mahalanobis distance–Understanding the math with examples (python) [Электронный ресурс]. Режим доступа: <https://www.machinelearningplus.com/statistics/mahalanobis-distance> (дата обращения: 09.11.2020).
2. Рыбников А. Технологический мониторинг бурения скважин в ПК DrillScene [Электронный ресурс]. Режим доступа: rozar.ru/wp-content/uploads/2017/06/6.Рыбников_Механический-мониторинг-бурения-скважин-в-ПК-DrillScene.pdf (дата обращения: 09.11.2020).
3. Алексеев А. Цифровое бурение [Электронный ресурс]. Режим доступа: <https://www.gazprom-neft.ru/press-center/sibneft-online/archive/2019-july-august/3406689/> (дата обращения: 09.11.2020).
4. Dursun S., Tuna T., Duman K., Kellog R.W. Real-time risk prediction during drilling operations. WO 2015060865A1 (Patent) 2015 [Электронный ресурс]. Режим доступа: <https://www.google.com/patents/WO2015060865A1?cl=en> Google Patents (дата обращения: 09.11.2020).
5. Shadizadeh S.R., Karimi F., Zoveidavianpoor M. Drilling stuck pipe prediction in Iranian oil fields: An artificial neural network approach // *Iranian Journal of Chemical Engineering*. 2010. Vol. 7. No. 4 (Autumn). P. 29–41.
6. Руководство пользователя. Программный комплекс «Инженерные расчеты строительства скважин». М.: Бурсофтпроект, 2017.
7. Van der Maaten L., Hinton G. Visualizing data using t-SNE // *Journal of Machine Learning Research*. 2008. Vol. 9. P. 2579–2605.
8. Университет ИТМО. Оценка качества в задачах классификации [Электронный ресурс]. Режим доступа: http://neerc.ifmo.ru/wiki/index.php?title=Оценка_качества_в_задачах_классификации (дата обращения: 09.11.2020).
9. Staudemeyer R.C., Morris E.R. Understanding LSTM–A tutorial into long short-term memory recurrent neural networks [Электронный ресурс]. Режим доступа: <https://arxiv.org/pdf/1909.09586.pdf> (дата обращения: 09.11.2020).
10. Breiman L., Cutler A. Random forests [Электронный ресурс]. Режим доступа: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (дата обращения: 09.11.2020).
11. Zharmagambetov A., Hada S.S., Carreira-Perpinan M.A., Gabidolla M. An experimental comparison of old and new decision tree algorithms. [Электронный ресурс]. Режим доступа: <https://arxiv.org/pdf/1911.03054.pdf> (дата обращения: 09.11.2020).
12. Ghosal I., Hooker G. Boosting random forests to reduce bias; One-step boosted forest and its variance estimate [Электронный ресурс]. Режим доступа: <https://arxiv.org/pdf/1803.08000.pdf> (дата обращения: 09.11.2020).
13. Brownlee J. Linear regression for machine learning [Электронный ресурс]. Режим доступа: <https://machinelearningmastery.com/linear-regression-for-machine-learning> (дата обращения: 09.11.2020).
14. Lundberg S., Lee S. A Unified approach to interpreting model predictions [Электронный ресурс]. Режим доступа: <https://arxiv.org/pdf/1705.07874.pdf> (дата обращения: 01.10.2020).

REFERENCES

- (1) Prabhakaran S. *Mahalanobis distance–Understanding the math with examples (python)*. Available from: <https://www.machinelearningplus.com/statistics/mahalanobis-distance> [Accessed: 9 November 2020].
- (2) Rybnikov A. *Monitoring of drilling process in DrillScene software package*. Available from: rozar.ru/wp-content/uploads/2017/06/6.Рыбников_Механический-мониторинг-бурения-скважин-в-ПК-DrillScene.pdf [Accessed: 9 November 2020]. (In Russian)
- (3) Alexeev A. *Digital drilling*. Available from: <https://www.gazprom-neft.ru/press-center/sibneft-online/archive/2019-july-august/3406689/> [Accessed: 9 November 2020]. (In Russian)
- (4) Dursun S, Tuna T, Duman K, Kellog RW. *Real-time risk prediction during drilling operations*. WO2015060865A1 (Patent) 2015.
- (5) Shadizadeh SR, Karimi F, Zoveidavianpoor M. Drilling stuck pipe prediction in Iranian oil fields: An artificial neural network approach. *Iranian Journal of Chemical Engineering*. 2010; 7(4): 29–41.
- (6) 000 Bursoftproekt. *Well Construction Engineering [Inzhenernye raschety stroitel'stva skvazhin] software package. User manual*. Moscow: Bursoftproekt; 2017. (In Russian)
- (7) Van der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*. 2008; 9: 2579–2605.
- (8) ITMO University. *Quality assessment in classification tasks*. Available from: http://neerc.ifmo.ru/wiki/index.php?title=Оценка_качества_в_задачах_классификации [Accessed: 9 November 2020]. (In Russian)
- (9) Staudemeyer RC, Morris ER. *Understanding LSTM–A tutorial into long short-term memory recurrent neural networks*. Available from: <https://arxiv.org/pdf/1909.09586.pdf> [Accessed: 9 November 2020].
- (10) Breiman L, Cutler A. *Random forests*. Available from: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm [Accessed: 9 November 2020].
- (11) Zharmagambetov A, Hada SS, Carreira-Perpinan MA, Gabidolla M. *An experimental comparison of old and new decision tree algorithms*. Available from: <https://arxiv.org/pdf/1911.03054.pdf> [Accessed: 9 November 2020].
- (12) Ghosal I, Hooker G. *Boosting random forests to reduce bias; One-step boosted forest and its variance estimate*. Available from: <https://arxiv.org/pdf/1803.08000.pdf> [Accessed: 9 November 2020].
- (13) Brownlee J. *Linear regression for machine learning*. Available from: <https://machinelearningmastery.com/linear-regression-for-machine-learning> [Accessed: 9 November 2020].
- (14) Lundberg S, Lee S. *A Unified approach to interpreting model predictions*. Available from: <https://arxiv.org/pdf/1705.07874.pdf> [Accessed: 1 October 2020].