

УДК 622.24.08:004.942.032.26

Нейросетевая модель прогнозирования скорости и режимов бурения скважин в сложнопостроенных коллекторах

Ю. Е. Катанов

Тюменский индустриальный университет, Тюмень, Россия
e-mail: katanov-juri@rambler.ru

Аннотация. Рассмотрена проблема, связанная с исследованием скоростей бурения скважин в сложнопостроенных коллекторах, решение которой представлено в виде нейросетевой модели, учитывающей структурные, геомеханические и технологические особенности системы «массив — скважина».

Представлена возможность прогнозирования способа бурения скважин с различными прочностными и структурно-литологическими характеристиками массива на базе нейросетевого моделирования.

Целью данного исследования является получение математических моделей для исследования вероятностно-статистических закономерностей процессов бурения скважин в условиях неопределенности.

Научной новизной выполненной работы является качественная и количественная оценки взаимовлияния геолого-технологических факторов на скорость бурения скважин; поиск оптимальных режимов бурения скважин в условиях сложнопостроенных коллекторов на базе математического моделирования.

Ключевые слова: нейросеть; подвыборка; персептрон; радиально-базисные функции; кросс-проверка; прогнозирование

Neural network model of the wells' drilling speed and modes predicting in complex reservoirs

Yury E. Katanov

Industrial University of Tyumen, Tyumen, Russia
e-mail: katanov-juri@rambler.ru

Abstract. The article considers the problem connected with the study of well drilling rates in complex reservoirs. Its solution is presented in the form of a neural network model that takes into account the structural, geomechanical and technological features of the «rock mass — well» system.

The possibility of predicting the well drilling method with different strength and structural-lithological characteristics of the massif, based on neural network modeling, is presented.

The purpose of this study is to obtain mathematical models for analysis of the probabilistic and statistical patterns of well drilling processes in conditions of uncertainty.

The scientific novelty of the work performed is the qualitative and quantitative assessment of the mutual influence of geological and technological factors on the well drilling rate; search for optimal well drilling modes in complex reservoirs on the basis of mathematical modeling.

Key words: neural network; subsampling; perceptron; radial-basis functions; cross-validation; forecasting

Введение

Для многих месторождений Среднего Приобья характерны терригенные отложения, представленные песчаниками, алевролитами, глинистыми породами с различными структурно-текстурными особенностями [1–3].

Для горных пород, представленных совокупностями линзовидных формаций, которые не имеют единой структурной формы, связанных с закономерностями формирования латеральных экранов, характерны различные тепловые аномалии и высокие пластовые давления.

Подобная корреляционная закономерность может быть охарактеризована как комплексирование результатов исследования зон трещиноватости и зон глубинных разломов, которые способствуют возникновению аномальных тепловых полей и разветвленных систем трещин, что значительно усложняет исследования закономерностей цифрового воссоздания геологических структур.

Матрица трещинно-поровых коллекторов практически непроницаема, а в межзерновой пористости песчаников (не более 2–3 %) проницаемость матрицы полностью будет отсутствовать [4, 5].

Трещинные типы коллекторов особенно развиты в толщах аргиллитов, что не позволит (практически полностью) их детально идентифицировать даже по данным геофизических исследований. Однако для песчаников и алевролитов баженовской и иных свит в комплексах переслаивания структур литологических типов допустима возможность идентификации неоднородных толщ, базирующаяся на вероятностно-статистической оценке изменения их структуры на этапах проектирования залежей углеводородов.

На месторождениях Восточной Сибири коллекторы трещинно-порового, трещинного и иных типов в толщах терригенных отложений представлены песчаниками и алевролитами с незначительными петрофизическими характеристиками, по причине присутствия в них развитых разнонаправленных систем трещин различной длины.

В условиях таких сложнопостроенных коллекторов требуется комплексный подход, включающий адекватную математическую базу для формирования оптимальной технологической схемы при проектировании бурения в условиях неопределенности поведения системы «массив — скважина».

Объект и методы исследования

Целью данного исследования является получение математических моделей для оценки механической скорости бурения в сложнопостроенных породах, а также алгоритмов исследования режимов бурения скважин в условиях неопределенности [6, 7].

Для написания научной работы были использованы исследования особенностей горных пород, механизмы построения нейросетевых моделей и способы их применения, представленные в работах Ю. Е. Катанова, А. К. Ягафарова, И. П. Попова и других отечественных и зарубежных ученых [8, 9].

Экспериментальная часть/постановка эксперимента

Нейронная сеть представляет собой математический инструмент, который способен моделировать любую непрерывную функцию $f(\vec{x}; W)$ (рис. 1).



Рис. 1. Схема формирования нейронной сети

Нейросети представлены самообучающимися системами (алгоритмами), которые позволяют идентифицировать огромное количество данных при реализации наиболее вероятных прогнозов для принятия решений в индивидуальных ситуациях.

Возможности обучения и самообучения являются одними из главных преимуществ нейросетей перед традиционными методами исследования различных закономерностей.

С математической точки зрения процесс обучения нейросетей представляется в виде многопараметрической задачи нелинейной оптимизации.

Как правило, нейросети используются для моделирования взаимосвязей, когда не известен точный вид данных связей между исследуемыми зависимыми (выходными) и независимыми (входными) переменными.

Основу нейросети составляют искусственные нейроны, которые позволяют на практике реализовать разнородные нелинейные функции для многих исследуемых переменных с привлечением любого количества скрытых нейронов:

- получение моделей классификации и распознавание образов;
- получение моделей принятия решений в условиях неопределенности;
- получение моделей кластеризации и поиск скрытых зависимостей;
- получение моделей прогнозирования;
- формирование алгоритмов снижения размерностей исследуемых систем и т. д.

Искусственные нейроны представлены совокупностями входных сигналов в виде вещественных чисел в отрезке $[0; 1]$, формирующих нелинейные функции, характеристиками которых являются весовые коэффициенты.

Целостная структура таких функций представлена диадой нелинейного преобразователя и сумматора.

Все нейроны обладают входными однонаправленными связями (синапсы), посредством которых происходит соединение с выходными сигналами синоптических сочетаний других нейронов (аксон).

Каждый синапс (отдельная связь) характеризуется весовой характеристикой текущей связи в виде взвешенных сумм всех сигналов, подаваемых на вход.

В результате идентификации всех нейросетевых формаций на выходе будет получено единственное решение, соответствующее наиболее вероятностному исходу.

Следует отметить, что в нейросетях реализуется принцип параллельной обработки сигналов для объединения большого числа нейронов в отдельные «слои», имеющие индивидуальные конфигурации и концепты взаимодействия по отношению к другим формациям (принцип послойного взаимодействия). Например, при интерпретации объединения геологических слоев в пачки, серии пачек, в массивы и т. д.

В качестве универсального механизма построения сетевой архитектуры многослойных нейросетей (задачи классификации и регрессии) можно использовать многослойные перцептроны с различными вариациями передаточных функций или обратными (перекрестными или последовательными) связями.

Стратегии построения нейросетей представлены пользовательской нейросетью (формируется только одна архитектура нейросети на усмотрение исследователя), автоматической нейросетью (формируются множества различных архитектур нейросетей, среди которых выбираются оптимальные с точки зрения показателей производительности), стратегией многократных подвыборок (бутстреп или случайные подвыборки, основанные на построении ансамблей нейросетей в рамках единой архитектуры для разных подвыборок).

Бутстреп в статистической обработке разнородных данных (Bootstrap) — это информационно-аналитический метод исследования вероятностных распределений информационных массивов, который основан на многократной генерации выборок (подвыборок) методом Монте-Карло на базе имеющейся исходной совокупности.

Интерпретация результатов исследования может быть реализована как для случаев увязки к отдельной нейросети и ансамблю нейросетей, так и по отдельным выборкам (контрольная, обучающая, тестовая).

Например, можно создать нейросеть для исследования геологотехнологических параметров конкретной скважины; исследовать структурно-деформационные характеристики массива при «его работе» на сжатие, растяжение и сдвиг (в соответствии с работой нагнетательных, добывающих, поисково-разведочных и пробных скважин).

Информационно-математическую основу любых нейросетей формируют:

- линейная модель $f(\vec{x}, W) = \sum_j x_j \cdot w_j$;
- обобщенная линейная модель $f(\vec{x}, W) = \varphi(\sum_j x_j \cdot w_j)$;
- нелинейная модель $f(\vec{x}, W) = \sigma(\sum_k x_k \cdot w_k(\vec{x}, W))$,

где \vec{x} — вектор всех входных сигналов нейронов; W — весовые коэффициенты исходных параметров; x_j — сетевые сигналы для линейной нейросетевой модели; w_j — сетевые веса для линейной нейросетевой модели; σ — значение ширины окна используемой функции (функций) построения нейросети; x_k — сетевые сигналы для нелинейной нейросетевой модели; w_k — сетевые веса для нелинейной нейросетевой модели; j, k — соответственно, количество компонент скрытых слоев для обобщенной линейной и нелинейной нейросетевых моделей.

Перед тем как создать нейросеть для исследования и идентификации оптимальных значений механической скорости и режимов бурения сложнопостроенных коллекторов, рассмотрим задачу прогнозирования временного ряда на базе модели АРППС (ARIMA) (модель проинтегрированного скользящего среднего и авторегрессии; AutoRegressive Integrated Moving Average), математическая основа которой представлена динамикой изменения значений механической скорости разведочного бурения на первые 200 метров проходки (рис. 2).

Для сглаживания исходных данных (минимизации пиковых значений на всем интервале в соответствии с выбранной шкалой) был использован спектральный анализ Фурье [10].

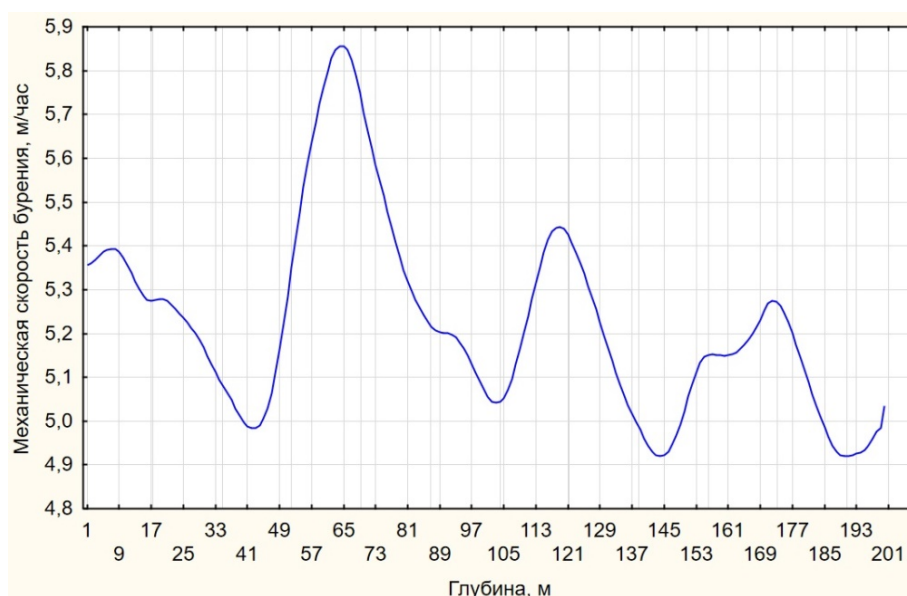


Рис. 2. Сглаженный временной ряд

На рисунке 2 прослеживаются явный полиномиальный тренд и не совсем явная периодичность, которая является базисной характеристикой временного ряда, благодаря которой можно осуществить поиск адекватной модели прогнозирования.

Периодичность не всегда можно «определить на глаз», поэтому можно построить периодограмму, наибольший пик которой будет соответствовать искомому периоду для идентификации значения временного лага (уровень влияния одного события на другие в течение некоторого временного интервала).

При невозможности идентификации временного лага задача прогнозирования на базе модели АРППС (ARIMA), а также при нейросетевом моделировании значительно усложняется, поскольку увеличивается вероятностный размах кросс-энтропии при выборе обучающих параметров системы «массив — скважина».

На рисунке 3 представлен прогноз значений механической скорости бурения, достоверность (вероятность такого исхода) которого равна 0,95916 при средневзвешенной ошибке прогноза 2,66 %.

Размах возможных значений механической скорости бурения с учетом интервалов допустимых значений равен (4,69; 5,09) м/ч.

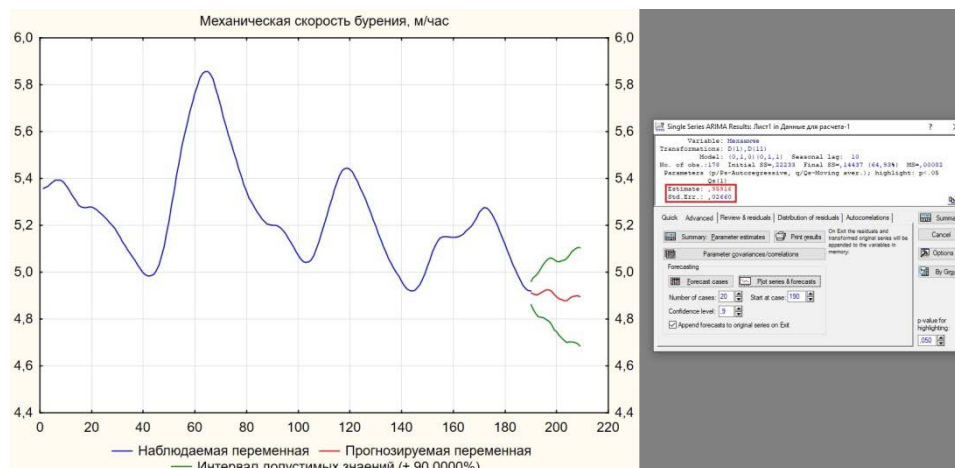


Рис. 3. Прогноз механической скорости бурения

Чтобы снизить такой уровень неопределенности, целесообразно создать альтернативный математический инструмент для исследования параметров бурения — нейросеть в задачах прогнозирования и классификации, выполненный, для примера, в пакете моделирования STATISTICA и содержащий в своей расчетной базе необходимые массивы данных [4, 7].

Для нейросетевой модели регрессия будет означать, что исследуемая переменная временного ряда является непрерывной.

В качестве начального условия, касающегося того факта, что не будут использованы определенные наблюдения из общей совокупности (для примера, последние 10 наблюдений), будет означать, что данные случаи будут оставлены для кросс-проверки (сквозная проверка по модели).

Всего в исходной базе данных было 200 случаев, следовательно, в задаваемом диапазоне для построения нейросети будет использовано с 1 по 190 случаев, без последних 10.

В качестве стратегии построения нейросети будет выбрана автоматическая нейронная сеть (АНС).

Для данной стратегии тестовые подвыборки задаваться не будут, поскольку уже было оставлено 10 последних наблюдений для кросс-проверки.

Для контрольной выборки, которая используется для остановки процесса обучения нейросети, будут взяты 30 % от всех наблюдений в произвольном порядке.

Это работает следующим образом: если фиксируется увеличение ошибки расчетов на контрольной выборке, то это свидетельствует о преоблада-

нии события переобучаемости нейросетевой модели и необходимости отклонения концепции ее дальнейшего обучения.

Как правило, необходимое количество наблюдений, подаваемых на вход в нейросеть, достаточных для ее адекватного обучения, будет определено размером окна нейросети.

Для автоматизированной стратегии исследования временных рядов использование моделей РБФ (RBF) (Радиальные базисные функции, Radial basis functions) будет нецелесообразным по причине их ориентирования на периодические ряды, которые не содержат глобальный тренд.

Поэтому основу случаев использования РБФ (RBF) составляют задачи интерполирования, когда исходные данные содержат множества пустых ячеек, которые целесообразно восстановить перед последующей их совместной обработкой с цельным массивом информации.

Нейросети, сформированные на РБФ (RBF), будут воспроизводить гауссовские поверхности откликов (результаты моделирования), поскольку они содержат промежуточные слои, состоящие из различных совокупностей радиальных элементов [11].

Поскольку функции откликов являются нелинейными, то использовать более одного промежуточного слоя в рамках модели восстановления произвольных нейросетевых зависимостей будет нецелесообразно.

Возникает вопрос, как корректно сформировать комбинации выходов для скрытых радиальных элементов нейросети и выбрать их необходимое количество, при дальнейшей идентификации наиболее оптимального выхода [12].

Для оптимального решения данного вопроса можно представить в виде линейных комбинаций исследуемые радиальные элементы (в виде взвешенных сумм гауссовых функций с понижениями по краям и с вершиной в центре), чтобы характеристики выходного слоя нейросети РБФ (RBF) содержали линейные функции активации [13].

Радиальные элементы задаются своими центром и «радиусом». Положения исследуемых точек в N-мерном пространстве будет определено N числовыми параметрами, число которых равняется количеству весовых характеристик у всех линейных элементов. Следовательно, координаты центров всех радиальных элементов будут тождественны их «весам». Радиусы (отклонения) исследуемых характеристик будут определены как «пороги».

Не следует путать понятия «весов» и «порогов» для радиальных и линейных элементов, поскольку радиальные «пороги» представлены отклонениями, радиальные «веса» — точками.

Однако для автоматизированной стратегии могут быть задействованы модели МЛП (MLP) (Многослойный перцептрон, Multilayered perceptron), поскольку данный механизм построения нейросетей является универсальным для любых задач, поэтому он будет задействован в данном исследовании.

В процессе линейного моделирования определение конфигураций нейросетей, предоставляющих возможность идентификации абсолютных минимумов для указанных интервалов допустимых ошибок, алгоритмически допустимо [14].

При использовании нелинейных возможностей нейросетей может быть сформирован инструментарий их корректировки с целью снижения уровня ошибок, когда нет абсолютной уверенности в том, что создание новых вариантов нейросетевых моделей позволит добиться еще меньших ошибок в исследованиях.

В подобных допущениях очень полезно понятие поверхности ошибок, суть которого сводится к следующему. Каждому из «весов» и «порогов» нейросети (при обозначении N как общего числа свободных параметров модели) ставится в соответствие одно измерение в формируемом многомерном пространстве.

Тогда $N+1$ измерение будет соответствовать ошибке нейросети, как значение-индикатор выхода за пределы области ее функционирования.

Поэтому для любых сочетаний исследуемых «весов» в $N+1$ -мерном пространстве представление в виде отдельных точек соответствующих величин ошибок нейросети позволит выделить их совокупность, называемую поверхностью ошибок.

Одной из главных задач корректного обучения нейросети является поиск на этом многомерном подмножестве самой низкой точки [15].

Для традиционных моделей (например, линейных), в качестве функции ошибок для которых выступает сумма квадратов (SOS — Sum of squares), формируемая поверхность ошибок геометрически будет представлена параболоидом (квадрикой) — некой гладкой поверхностью, которая похожа на часть поверхности сферы, с единственным минимумом. Для таких ситуаций локализовать этот минимум будет достаточно просто.

В случае использования нейросетей, поверхность ошибок будет представлена усложненным геометрическим строением, которое может содержать лишь локальные минимумы (отдельные точки, расположенные ниже остальных в индивидуальной окрестности, но находящиеся выше чем глобальный минимум), подмножества седловых точек, различные плоские участки, узкие и длинные геометрические «овраги» [16].

Аналитико-графическими средствами нельзя будет определить положение такого глобального минимума на поверхности ошибок, поэтому один из приоритетов адекватного обучения нейросети, по сути дела, и будет заключаться в исследовании поверхности ошибок.

При случайной конфигурации начальных «весов» и «порогов» (то есть если взять случайную точку на поверхности ошибок) совершенствующийся алгоритм обучения постепенно отыщет глобальный минимум. Как правило, для этого вычисляется наклон (градиент) поверхности ошибок в данной случайной точке, а затем полученная информация будет использована для дальнейшего «продвижения вниз» относительно геометрического «склона».

В итоге алгоритм остановится в некоторой нижней точке, которая, возможно, окажется лишь локальным минимумом (в лучшем случае — глобальным минимумом).

Для персептронов индивидуально установленный интервал скрытых нейронов (для данного примера) будет содержать их минимальное и максимальное количество от 5 до 20, а для обучения сети будут использованы 50 слоев; в качестве функции ошибки будет использована Sum of squares

(SOS — Сумма квадратов), потому что по своей сути любой временной ряд представляет собой задачу регрессии.

Для входных нейронов в качестве функций активации моделей МЛП (MLP), соответственно, будут заданы Tanh, Logistic, Sine, Exponential (Гиперболическая, Логистическая, Синусоидальная, Экспоненциальная), а для выходных нейронов — Tanh, Logistic, Sine, Identity, Exponential (Гиперболическая, Логистическая, Синусоидальная, Тождественная, Экспоненциальная).

Затухание нейросети использоваться не будет для данной задачи.

В качестве установок для рандомизации можно выбрать начальное значение подвыборок, что приведет к тому, что для всех сетей будут генерироваться одни и те же обучающие контрольные подвыборки.

Этот прием полезен в тех случаях, когда требуется посмотреть на качество модели именно на одних и тех же подвыборках. Но поскольку для данного примера используются разные подвыборки, то данный прием не требуется [17, 18].

Итог обучения сети представлен на рисунке 4.

Слева направо: Первый столбец — Net name (Архитектура сети), например, MLP-10-8-1 расшифровывается как — используется модель многослойного персептрона МЛП (MLP), имеющая 10 входов, 8 скрытых нейронов и 1 выход.

Второй столбец — Training perf. (Производительность обучения), характеризуется величиной корреляции между искомым рядом и предсказанным, соответственно, на каждой подвыборке (чем значение больше, тем лучше сформированная сеть).

Третий столбец — Test perf. (Контрольная производительность) — охарактеризован возможной поддержкой корреляций на контрольных (обучающих) подвыборках и при минимизации между ними резких отличий, что позволит предположить равномерную работу нейромодели как на обучении, так и на контроле.

Четвертый столбец пропущен, так как не было проверочных этапов (заранее были оставлены 10 последних наблюдений для кросс-проверки).

Пятый и шестой столбцы — Training error (Ошибка обучения) и Test error (Контрольная ошибка), соответственно, численные значения функционала ошибки как на обучении, так и на контроле.

Седьмой столбец пропущен.

Восьмой столбец — Training algorithm (Алгоритм обучения) и итерация (значение, указанное после аббревиатуры алгоритма), на которой этот алгоритм был остановлен.

Девятый, десятый и одиннадцатый столбцы — соответственно, Error function (Функция ошибки) (SOS — Сумма квадратов), Hidden activation (Функция активизации скрытых нейронов) и Output activation (Функция активизации выходных нейронов).

Для визуальной оценки качества моделей можно воспользоваться диаграммой рассеяния: поле X-axis определено целевой зависимостью (Target); поле Y-axis — выходной зависимостью (Output) (рис. 5).

STATISTICA - [Данные для расчета-2* - Summary of active networks (Лист1 in Данные для расчета-2)]

File Edit View Format Statistics Data Mining PROCEED Graphs Tools Workbook Scorecard

Neural Networks Machine Learning MARS Boosted Random Forests

Workspaces Optimizations Feature Selection

Association Rules Link Analysis Rule Extraction

Rapid Deployment Goodness of Fit

Deployment

Tools

Summary of active networks (Лист1 in Данные для расчета-2)

Include cases: 1130

Index	Net name	Training perf.	Test perf.	Validation perf.	Training error	Test error	Validation error	Training algorithm	Error function	Hidden activation	Output activation
1	MLP 10-8-1	0.990032	0.987522	0.987522	0.00577	0.00683		BFGS 10	SOS	Sine	Identity
2	MLP 10-6-1	0.991193	0.98035	0.98035	0.00503	0.00660		BFGS 25	SOS	Tanh	Exponential
3	MLP 10-8-1	0.994631	0.98300	0.98300	0.00301	0.00568		BFGS 140	SOS	Exponential	Exponential
4	MLP 10-4-1	0.991390	0.985981	0.985981	0.00489	0.00788		BFGS 24	SOS	Tanh	Logistic
5	MLP 10-5-1	0.991571	0.986415	0.986415	0.00481	0.00744		BFGS 55	SOS	Logistic	Tanh
6	MLP 10-6-1	0.993632	0.986133	0.986133	0.00360	0.00780		BFGS 90	SOS	Exponential	Exponential
7	MLP 10-8-1	0.991819	0.986433	0.986433	0.00465	0.00744		BFGS 36	SOS	Tanh	Tanh
8	MLP 10-3-1	0.990245	0.986512	0.986512	0.00461	0.00775		BFGS 23	SOS	Logistic	Exponential
9	MLP 10-8-1	0.990145	0.987058	0.987058	0.00561	0.00698		BFGS 10	SOS	Identity	Identity
10	MLP 10-2-1	0.990242	0.987074	0.987074	0.00568	0.00709		BFGS 15	SOS	Identity	Identity

SANN - Results: Лист1 in Данные для расчета-2

Active neural networks

Net ID	Net name	Training perf.	Test perf.	Validation perf.	Algorithm	Error funct.
1	MLP 10-8-1	0.990032	0.987522	0.987522	BFGS 10	SOS
2	MLP 10-6-1	0.991193	0.98035	0.98035	BFGS 25	SOS
3	MLP 10-8-1	0.994631	0.98300	0.98300	BFGS 140	SOS
4	MLP 10-4-1	0.991390	0.985981	0.985981	BFGS 24	SOS

Select/Unselect active networks

Delete networks

Build models with CNN

Build models with ANS

Build models with Subsampling

Predictions | Graphs | Details | Time series | Custom predictions

Predictions spreadsheet

Predictions type

Standalone

Ensemble

Standalone and ensemble

Include

Inputs

Targets

Output

Residuals

Variables

Confidence intervals

Save networks

Cancel

Options

Samples

Train

Test

Validation

Missing

Рис. 4. Характеристики сформированных нейросетей

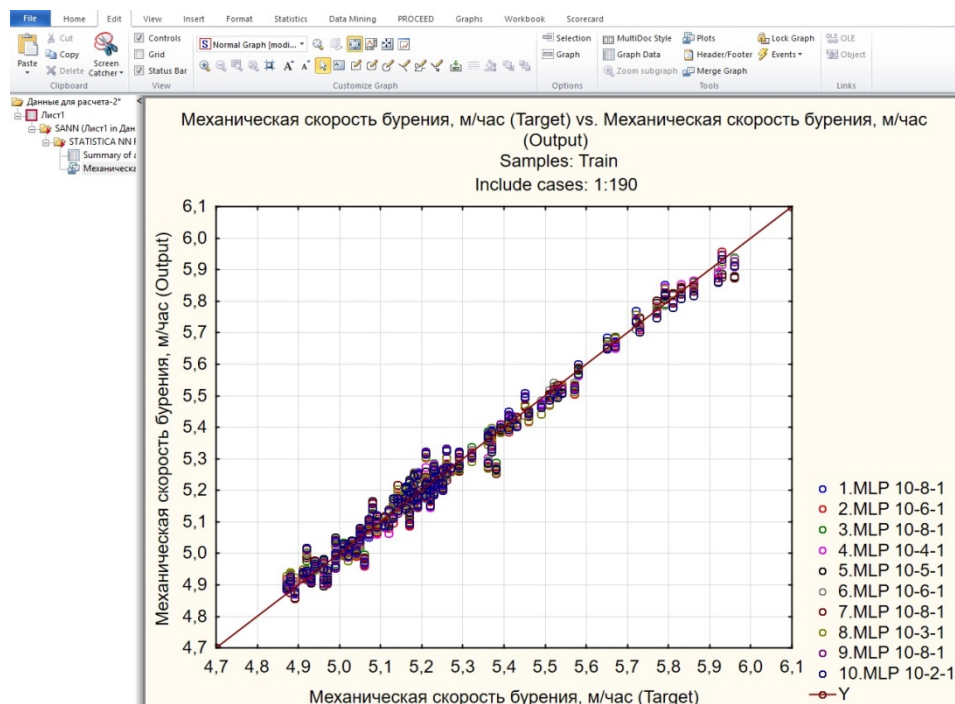


Рис. 5. Диаграмма рассеяния

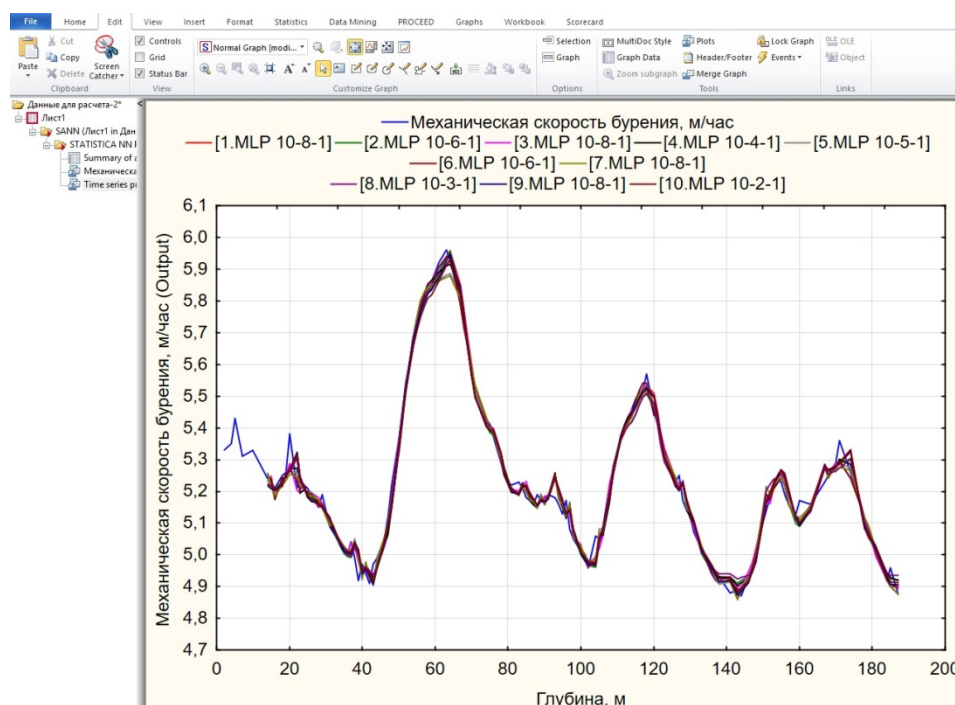


Рис. 6. Сопоставление прогнозов по нейромоделям и искомого ряда значений механической скорости бурения

Чем ближе исследуемые точки будут к выделенной прямой линии — тем лучше будет качество построенных нейросетей.

На рисунке 6 искомый ряд представлен «синим цветом», а предсказанные ряды для каждой нейромодели — другой палитрой (ориентироваться по легенде сверху).

Для детальной интерпретации таких преобразований используется проекция временного ряда, относительно которой каждое значение прогнозного ряда будет построено с учетом предыдущих значений исходного ряда (см. рис. 6).

Пусть не смущает относительная хаотичность линий — если обратить внимание на шкалу, видно, что значения величин сильно не отличаются друг от друга (см. рис. 6).

Чтобы визуально посмотреть, как эти проекции ложатся на весь временной ряд, можно начинать построение проекционной картины сразу с 1 наблюдения до 200, и далее, на 20 шагов вперед (1 шаг равен 1 метру) до отметки 220 метров (рис. 7).

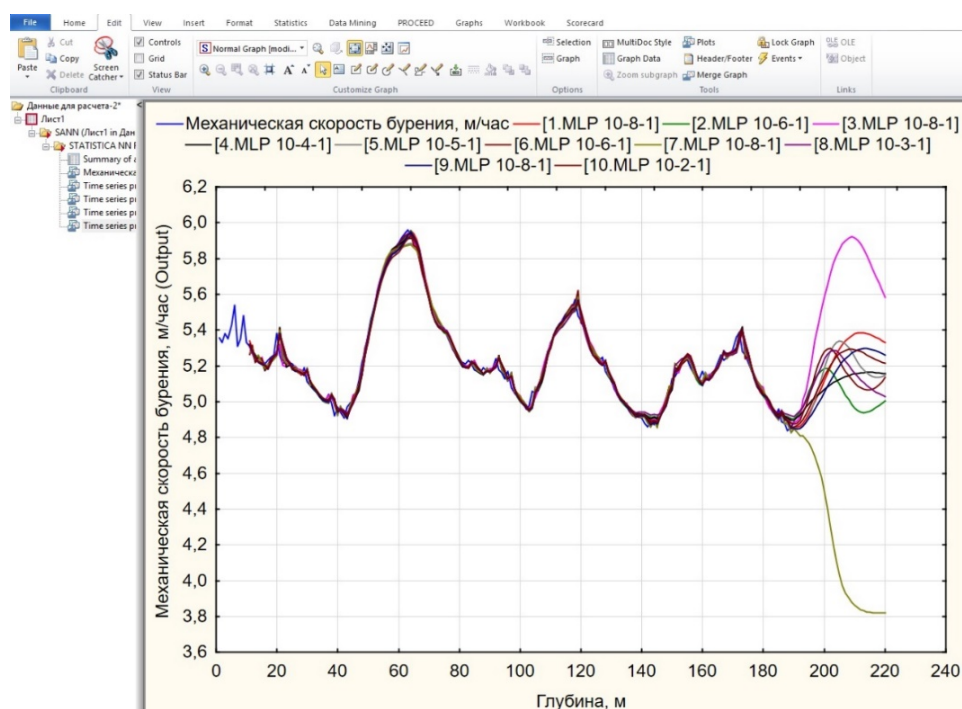


Рис. 7. График проекции с прогнозом на 20 метров вперед

Использовать все построенные нейромодели нет смысла. Чтобы определиться с выбором, какие из них удалить, можно посмотреть, как они проявят свою робастность (устойчивость) на большее число шагов вперед.

Из рисунка 7 видно, что есть модели, которые вообще не улавливают зависимость и их можно смело удалять — ориентируясь по легенде, это модели 3 и 7.

Из оставшихся моделей выбираем те (или ту), которые имеют минимальную ошибку прогноза и наибольшее значение производительности.

Наиболее оптимальными будут нейромодели 4 и 5.

Следует отметить, что цветовая палитра в легенде (при отображении результатов) изменяется по мере удаления не информативных сетей из общего списка.

Поэтому цвет ряда для модели 4 изменился на «красный», а для модели 5 на «зеленый» (рис. 8). Итоговый прогноз будет реализован относительно нейросетевой модели 5.

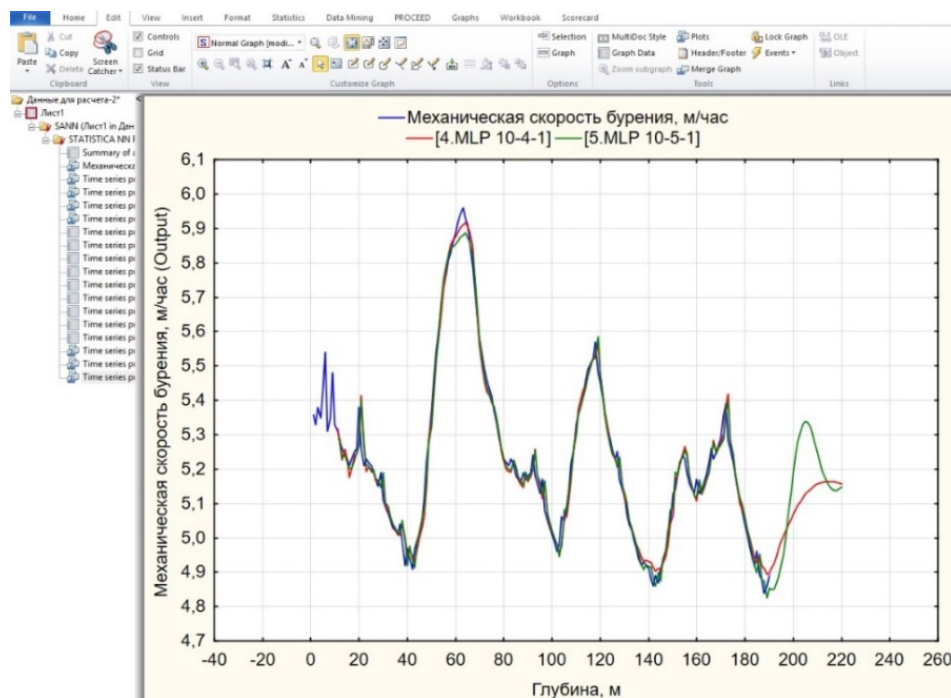


Рис. 8. Прогноз механической скорости бурения на базе нейромоделей

Если исходный ряд данных является сложным для исследования, то сразу ожидать идеальную нейромодель не нужно. Это достигается постепенно, по принципу от простого к сложному.

Далее можно анализировать качество сети.

Сравним прогноз построенной нейросети с тестовой выборкой, размер которой увеличим до количества значений исходного ряда (см. рис. 8).

Полученные ошибки прогноза, варьирующиеся в интервале (0,49–0,62 %), значительно ниже, чем вероятные ошибки прогноза, полученные при привлечении технологии прогнозирования на базе модели АРППС (ARIMA) (см. рис. 3).

Интерпретируя математическую основу нейросетей, можно утверждать, что результаты прогнозирования на базе временных рядов будут идентичны результатам многомерной регрессии в рамках нейромоделей, но соответствующие им области допустимых решений будут различными.

Процедуру поиска результатов регрессии целесообразно реализовать с учетом классической схемы построения математических моделей: при наличии адекватного базиса концептуальных и информационных моделей, далее, после привлечения необходимых численных методов начинать с

простых математических моделей, постепенно увеличивать их сложность посредством привлечения дополнительных условий до тех пор, пока качество построенных нейромоделей не станет удовлетворительным.

Теперь проведем классификацию временных рядов на базе нейромоделей для дополнения к уже выполненному исследованию.

Пусть исходные данные бурения содержат информацию о трех классах исследуемых объектов (Лопастные, Шарошечные, Алмазные долота), каждый из которых характеризуется 7 признаками (Диаметр долота, Диаметр УБТ (Утяжеленная бурильная труба), Расход промывочной жидкости, Средневзвешенная плотность пород, Осевая нагрузка, Механическая скорость бурения, Частота вращения).

Для задач классификации вводится еще одна переменная кодов, соответствующая настройке параметров обучения — для контрольной (Train) и обучающей (Select) выборки. Для классификации будут использованы три стратегии построения нейросетей: пользовательская нейросеть (ПНС), автоматическая нейросеть (АНС) и методы многократных подвыборок.

Какие из наборов данных будут обучающими подвыборками, а какие тестовыми, исследователь сам для себя решает при формировании исходных данных.

По стратегии Subsampling (Создание подвыборок), для примера будут построены 10 сетей. В задачах классификации, как правило, используются Logistic, Tanh (соответственно, логистические и гиперболические функции), как на Hidden neurons (Скрытые нейроны), так и на Output neurons (Выходные нейроны).

Процедура Weight decay (Затухание) будет использована для скрытого слоя нейронов посредством применения процедуры регуляризации (при внесении скрытых слоев). Для пользовательской нейросети можно выбрать построение с помощью РБФ.

Для данного примера количество сетей для обучения будет 5, а количество скрытых нейронов укажем равным 6 (выбор данного числа обусловлен тем, что имеется 6 классов в данной задаче классификации, поэтому минимальное их число должно быть не меньше числа классов).

Таким образом, суммарно по всем стратегиям было построено 22 модели, в которых указаны их архитектура, производительность на всех подвыборках, алгоритм обучения, функции ошибки и функции активации скрытых и выходных нейронов.

Также построена матрица ошибок по каждой сети, в которой будет указано число правильно (Correct) и неправильно (Incorrect) классифицированных наблюдений из общего их количества (Total) по каждой модели и каждому классу. В итоге необходимо будет выбрать те модели (из общего списка), которые имеют наибольшую производительность по всем подвыборкам (рис. 9).

Под производительностью модели классификации понимается процент корректного отнесения значений наблюдаемой переменной к значениям прогнозной переменной. Чем ближе величина производительности к 100 %, тем робастнее (устойчивее) построенная модель классифицирует исходные данные.

Для первой построенной нейросети с архитектурой MLP 7-12-3 (архитектура многослойного персептрона, 7 входов, 12 скрытых нейронов, 3 выхода) характерны следующие результаты.



Первый столбец Sample (Выборка) — тип выборки (обучающая или контрольная). Далее, слева направо распределены столбцы всех обрабатываемых переменных на входе (Input); после столбец целевого значения (Target) класса «Тип долота» и столбец выходного значения (Output), предсказанного сетью для данного класса; дальше представлены столбцы с вероятностями принадлежности наблюдений целевого класса к одному из трех (потому что было три класса переменных на входе — Алмазные, Лопастные и Шарошечные долота). Аналогично представлены результаты классификации для остальных нейросетей.

Принцип отнесения наблюдаемых значений к конкретному классу базируется на следующем: тот класс, которому соответствует наибольшая вероятность принадлежности, будет маркирован нейросистемой как исходный выходной класс. Значения, выделенные «красным цветом», указывают на ошибку отнесения к конкретному классу. В качестве оптимальной нейромодели классификации будет выбрана модель 20.

Построим лифтовые карты, которые позволяют как-то охарактеризовать качество классификации (рис. 10).

Принцип построения данных карт достаточно простой. Происходит сортировка всех наблюдений по вероятности принадлежности, например, к «шарошечным долотам». Далее выбирается количество наблюдений из сортированного списка в процентах (Percentile, условно ось X).

По карте видно, что первые 30 % наблюдений составляют около 70 % всех точек (условно ось Y), которые принадлежат «шарошечным долотам». При 30 % происходит охват всего множества точек, принадлежащих «шарошечным долотам».

Идеальным случаем для лифтовых карт (карт выигрышей) будут ситуации, когда убывающая кривая по построенной модели имеет только один явный угол к основной функции (чтобы было условное сочетание двух прямых). А в данной модели присутствуют 4 таких угла (см. рис. 10).

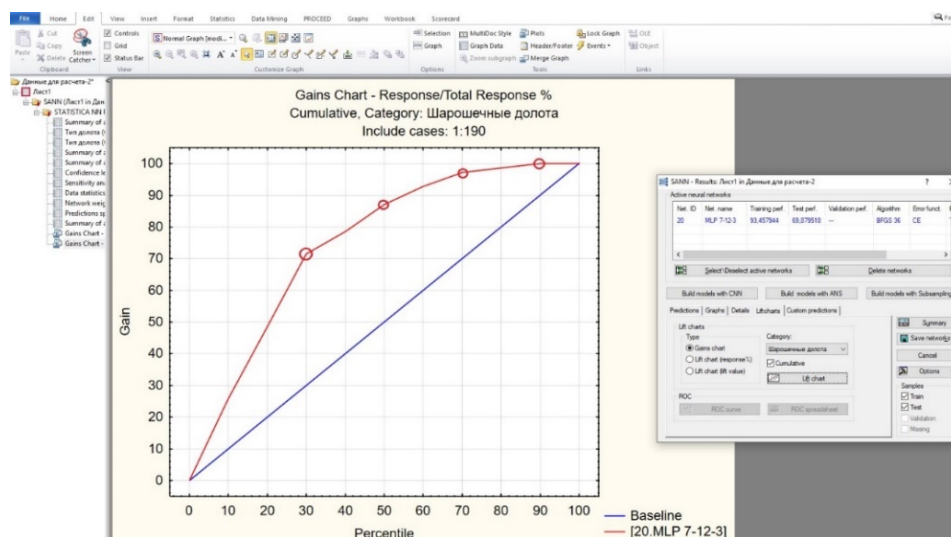


Рис. 10. Лифтовая карта

Лифтовая карта с откликом (Lift chart (response %)) представляется в следующем виде (рис. 11).

Относительно данного типа карты сортируются все наблюдения, и она показывает количество (процент) правильной классификации наблюдений.

Исходя из полученной зависимости, если возьмем 10 % отсортированных наблюдений (первая выделенная точка на оси X), то в них содержатся только элементы, принадлежащие «шарошечным долотам» (95 % — первая точка на оси Y). Если берем больше элементов (перемещаемся по оси X вправо), то накапливаются некоторые ошибки (спад кривой) и, соответственно, появляются элементы, которые не принадлежат к данному классу (снижение % по оси Y в соответствующих точках) [19, 20].

Данная кривая убывает до тех пор, пока не будут взяты все имеющиеся элементы (100 %). В данном примере значение 10 (условно ось X) формируется как отношение всех элементов «шарошечных долот» к элементам всей выборки. В идеальном случае для лифтовой карты с откликом убывающая кривая по построенной модели должна быть прямой [21, 22].

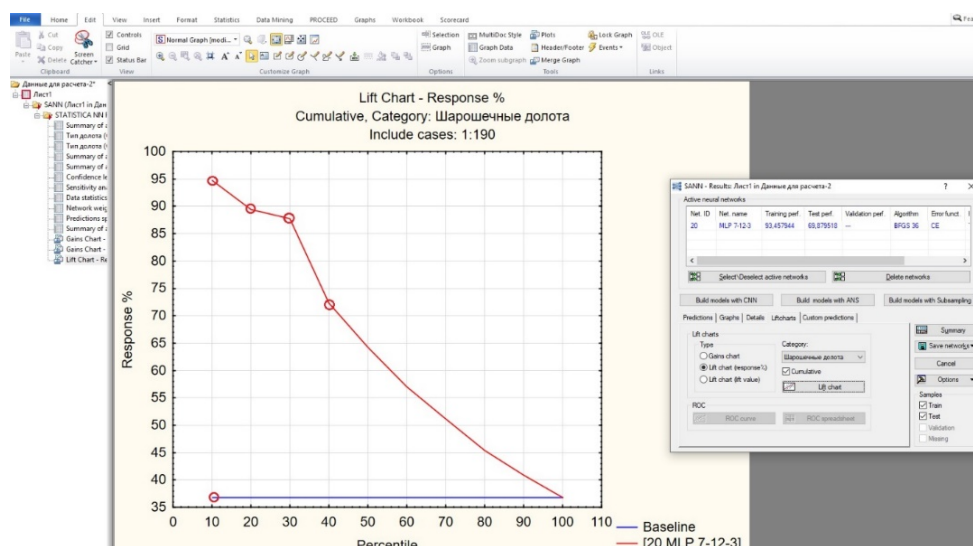


Рис. 11. Лифтовая карта с откликом

На основании выполненного нейросетевого моделирования был сформирован композиционный план, учитывающий полученные прогнозные значения механической скорости бурения на 20 метров глубже последней исследуемой отметки массива, на основании которого были выделены интервалы бурения и соответствующие им прогнозные значения скоростей и способов бурения при учете физико-механических характеристик пород и технологических особенностей возможного бурильного оборудования (рис. 12).

Итогом выполненного нейросетевого моделирования является разделение исследуемого бурением глубинного интервала на две стадии:

- для интервала бурения 200–214 метров величина средней механической скорости будет соответствовать средневзвешенному арифметическому значению (5,15; 5,34) — 5,27 м/ч, преимущественно для роторного

способа бурения при величине оборотов ротора в интервале 40–55 об/мин. Величина рекомендуемой производительности насосов и соответствующей нагрузки варьирует в пределах 40–46 л/с и 2–9 тонн;

- для интервала бурения 200–214 метров величина средней механической скорости будет соответствовать средневзвешенному арифметическому значению (5,14; 5,21) — 5,18 м/ч, преимущественно для турбинного способа бурения с рекомендуемой нагрузкой 2–9 тонн.

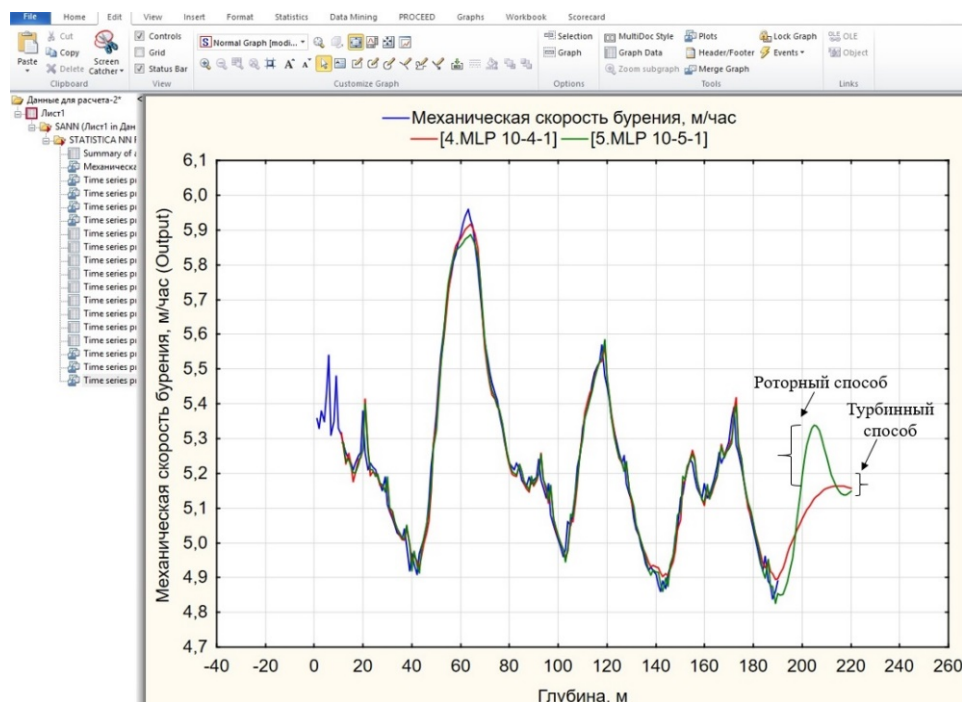


Рис. 12. Результаты нейросетевого моделирования

Результаты и обсуждение

Установлены закономерности изменения скорости бурения в сложнопостроенных коллекторах при различных технологических нагрузках. Предложена нейросетевая модель исследования способа бурения скважин, учитывающая структурные особенности геологического массива.

Результаты исследований были представлены на семинарах кафедр геологии месторождений нефти и газа, бурения нефтяных и газовых скважин, разработки и эксплуатации нефтяных и газовых месторождений, а также на научно-практических конференциях Тюменского индустриального университета.

Выводы

- На основании использования модели прогноза временного ряда АРППС (ARIMA) размах возможных значений механической скорости бурения 20 метров условной скважины, с учетом интервалов допустимых значений, равен (4,69; 5,09) м/ч при средневзвешенной ошибке прогноза 2,66 %.

- На основании нейросетевого моделирования при использовании механизмов построения нейросети, в частности, многослойного персептрона и радиально-базисных функций, прогнозная скорость бурения 20 метров условной скважины изменяется в интервале (4,85; 5,34) м/ч, с ее уточнением при смене способа бурения, при средневзвешенной ошибке прогноза 0,62 %.

В качестве рекомендации дальнейшего исследования можно предложить реализовать иерархическое комплексирование построенных нейросетевых моделей и алгоритмов решений с методами машинного обучения, что позволит включить в общий математический базис данные технологии, а также методы интеллектуального анализа геолого-промысловых и технологических данных.

Полученные аппроксимационные закономерности рекомендуется использовать при углубленном исследовании структурных особенностей массивов осадочного происхождения для дальнейшего прогнозирования скоростных характеристик бура в условиях неопределенности.

Библиографический список

1. Басарыгин, Ю. М. Технология бурения нефтяных и газовых скважин : учебник для вузов / Ю. М. Басарыгин, А. И. Булатов, Ю. М. Проселков. – Москва : Недра, 2001. – 679 с. – (Высшее образование). – Текст : непосредственный.
2. Булатов, А. И. Техника и технология бурения нефтяных и газовых скважин : учебник для вузов / А. И. Булатов, Ю. М. Проселков, С. А. Шаманов. – Москва : Недра, 2003. – 1007 с. – (Высшее образование). – Текст : непосредственный.
3. Балденко, Д. Ф. Винтовые забойные двигатели = Positive Displacement Downhole Motors : справочное пособие / Д. Ф. Балденко, Ф. Д. Балденко, А. Н. Гноевых. – Москва : Недра, 1999. – 375 с. – Текст : непосредственный.
4. Катанов, Ю. Е. Механизмы и принципы моделирования деформационно-пространственной неустойчивости горных пород / Ю. Е. Катанов. – Текст : непосредственный // Строительство нефтяных и газовых скважин на суше и на море. – 2014. – № 11. – С. 19–23.
5. Катанов, Ю. Е. Обоснование параметров горных пород (неоднородных сред) при моделировании / Ю. Е. Катанов, Н. В. Григорьев. – Текст : непосредственный // Международный научно-исследовательский журнал. – 2014. – № 8–1(27). – С. 18–21.
6. Иващенко, В. В. Прогнозирование продуктивных характеристик пород-коллекторов в условиях неопределенности / В. В. Иващенко, Ю. Е. Катанов. – Текст : непосредственный // Новая наука : Проблемы и перспективы. – 2016. – № 121–3. – С. 31–33.
7. Катанов, Ю. Е. Создание линейного симулятора для прогнозирования технологического процесса / Ю. Е. Катанов. – Текст : непосредственный // Известия высших учебных заведений. Нефть и газ. – 2012. – № 1. – С. 112–116.
8. Петухов, А. В. Теория и методология изучения структурно-пространственной зональности трещинных коллекторов нефти и газа / А. В. Петухов. – Ухта : Ухтинский государственный технический университет, 2002. – 276 с. – Текст : непосредственный.

9. Технологии повышения продуктивности скважин и воздействия на залежи углеводородов на месторождениях Западной Сибири : монография / А. К. Ягафаров, И. И. Клещенко, И. П. Попов [и др.] ; Министерство образования и науки Российской Федерации, Тюменский индустриальный университет. – Тюмень : ТИУ, 2017. – 204 с. – Текст : непосредственный.
10. Катанов, Ю. Е. Основы теории управления : учебное пособие / Ю. Е. Катанов ; Министерство науки и высшего образования Российской Федерации, Тюменский индустриальный университет. – Тюмень : ТИУ, 2019. – 170 с. – Текст : непосредственный.
11. Katanov, Yu. E. Geological and mathematical analogy of reservoir and polymer structures / Yu. E. Katanov, V. M. Aleksandrov, A. K. Yagafarov. – Текст : непосредственный // TEST Engineering & Management. – 2020. – P. 6977–6991.
12. Murphy, K. P. Machine Learning : A Probabilistic Perspective / K. P. Murphy. – MIT Press, Cambridge, MA, USA, 2012. – 1070 p. – Текст : непосредственный.
13. Morin, F. Hierarchical Probabilistic Neural Network Language Model / F. Morin, Y. Bengio. – Текст : непосредственный // Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics. – 2005. – P. 246–252.
14. Nowlan, S. J. Simplifying Neural Networks by Soft Weight-Sharing / S. J. Nowlan, G. E. Hinton. – DOI 10.1162/neco.1992.4.4.473. – Текст : непосредственный // Neural Computation. – 1992. – Vol. 4, Issue 4. – P. 473–493.
15. Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks / M. Oquab, L. Bottou, I. Laptev, J. Sivic. – Текст : непосредственный // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2014. – P. 1717–1724.
16. Saxe, A. M. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks / A. M. Saxe, J. L. McClelland, S. Ganguli. – Текст : электронный. – URL: <https://arxiv.org/abs/1312.6120>.
17. Катанов, Ю. Е. Компьютерные технологии / Ю. Е. Катанов ; Министерство науки и высшего образования Российской Федерации, Тюменский индустриальный университет. – Тюмень : ТИУ, 2020. – Ч. 1: Геология, геофизика, гидрогеология. – 180 с.
18. Катанов, Ю. Е. Разработка проекта информационного обеспечения «INTELLPRO» / Ю. Е. Катанов. – Текст : непосредственный // Новые информационные технологии в нефтегазовой отрасли и образовании : материалы V Всероссийской научно-технической конференции с международным участием (01–02 ноября 2012 года) / Отв. ред. О. Н. Кузяков – Тюмень : Тюменский индустриальный университет, 2012. – С. 20–25.
19. Меньшиков, А. И. Исследование оптимальных режимов работы шарошечных долот в условиях неопределенности / А. И. Меньшиков. – Текст : непосредственный // Наука в современном обществе : закономерности и тенденции развития : сборник трудов Международной научно-практической конференции, (Стерлитамак, 4 апреля 2019 года). – Уфа : ООО «Омега Сайнс», 2019. – С. 48–54.
20. Меньшиков, А. И. Кластеризации параметров бурения / А. И. Меньшиков. – Текст : непосредственный // Технологическое развитие науки : тенденции, проблемы и перспективы : сборник статей по итогам Международной научно-практической конференции (Стерлитамак, 8 июня 2019 года). – Уфа : ООО «Аэтерна», 2019. – С. 30–35.

21. Меньшиков, А. И. Исследование режимов работы шарошечных долот на основе деревьев решений / А. И. Меньшиков. – Текст : непосредственный // Информационные технологии как основа прогрессивных научных исследований : сборник трудов Международной научно-практической конференции (Пермь, 25 мая 2020 года). – Уфа : ООО «Аэтерна», 2020. – С. 78–83.

22. Меньшиков, А. И. Прогнозирование механической скорости бурения массива шарошечными долотами на основе нейросетевого моделирования / А. И. Меньшиков. – Текст : непосредственный // Информационные технологии как основа прогрессивных научных исследований : сборник трудов Международной научно-практической конференции (Пермь, 25 мая 2020 года). – Уфа : ООО «Аэтерна», 2020. – С. 83–87.

References

1. Basarygin, Yu. M., Bulatov, A. I., & Proselkov, Yu. M. (2001). Tekhnologiya bureniya neftyanykh i gazovykh skvazhin. Moscow, Nedra Publ., 679 p. (In Russian).
2. Bulatov, A. I., Proselkov, Yu. M., & Shamanov, S. A. (2003). Tekhnika i tekhnologiya bureniya neftyanykh i gazovykh skvazhin. Moscow, Nedra Publ., 1007 p. (In Russian).
3. Baldenko, D. F., Baldenko, F. D., & Gnoevykh, A. N. (1999). Positive Displacement Downhole Motors. Moscow, Nedra Publ., 375 p. (In Russian).
4. Katanov, Yu. E. (2014). Mechanisms and principals of modeling of rocks deformational-spatial instability. Construction of Oil and Gas Wells on Land and Sea, (11), pp. 19-23. (In Russian).
5. Katanov, Yu. E., & Grigoriev, N. V. (2014). Substantiation of the parameters of rocks (heterogeneity environments) when modeling. International research journal, (8-1(27)), pp. 18-21 (In Russian).
6. Ivashchenko, V. V., & Katanov, Yu. E. (2016). Prognozirovaniye produktivnykh kharakteristik porod-kollektorov v usloviyakh neopredelennosti. Novaya nauka: Problemy i perspektivy, (121-3), pp. 31-33.
7. Katanov, Yu. E. (2012). Creation of the linear simulator for prediction of the technological process. Higher Educational Institutions News. Neft' I Gas, (1), pp. 112-116. (In Russian).
8. Petukhov, A. V. (2002). Teoriya i metodologiya izucheniya strukturno-prostranstvennoy zonal'nosti treshchinnykh kollektorov nefi i gaza. Ukhta, Ukhta State Technical University Publ., 276 p. (In Russian).
9. Yagafarov, A. K., Kleshchenko, I. I., Popov, I. P., Vaganov, Yu. V., Katanov, Yu. E., Popova, Zh. S., & Savina, M. E. (2017). Tekhnologii povysheniya produktivnosti skvazhin i vozdeystviya na zalezhi uglevodorodov na mestorozhdeniyakh Zapadnoy Sibiri. Tyumen, Industrial University of Tyumen Publ., 204 p. (In Russian).
10. Katanov, Yu. E. (2019). Osnovy teorii upravleniya. Tyumen, Industrial University of Tyumen Publ., 170 p. (In Russian).
11. Katanov, Yu. E., Aleksandrov, V. M., & Yagafarov, A. K. (2020). Geological and mathematical analogy of reservoir and polymer structures. TEST Engineering & Management, pp. 6977-6991. ((n English).
12. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press, Cambridge, MA, USA, 1070 p. (In English).
13. Morin, F., & Bengio, Y. (2005). Hierarchical Probabilistic Neural Network Language Model. Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, pp. 246-252. (In English).

14. Nowlan, S. J., & Hinton, G. E. (1992). Simplifying Neural Networks by Soft Weight-Sharing. *Neural Computation*, 4(4), pp. 473-493. (In English). DOI: 10.1162/neco.1992.4.4.473
15. Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1717-1724. (In English).
16. Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. (In English). Available at: <https://arxiv.org/abs/1312.6120>.
17. Katanov, Yu. E. (2020). *Komp'yuternye tekhnologii. Chast' 1. Geologiya, geofizika, gidrogeologiya*. Tyumen, Industrial University of Tyumen Publ., 180 p. (In English).
18. Katanov, Yu. E. (2012). *Razrabotka proekta informatsionnogo obespecheniya "INTELLPRO". Novye informatsionnye tekhnologii v neftegazovoy otrasli i obrazovanii: materialy V Vserossiyskoy nauchno-tekhnicheskoy konferentsii s mezhdunarodnym uchastiem*, November, 1-2. Tyumen, Industrial University of Tyumen Publ., pp. 20-25. (In Russian).
19. Men'shikov, A. I. (2019). *Issledovanie optimal'nykh rezhimov raboty sharoshechnykh dolot v usloviyakh neopredelennosti*. Nauka v sovremennom obshchestve: zakonomernosti i tendentsii razvitiya: sbornik trudov Mezhdunarodnoy nauchno-prakticheskoy konferentsii, (Sterlitamak, April, 4). Ufa, Omega Sayns LLC Publ., pp. 48-54. (In Russian).
20. Men'shikov, A. I. (2019). *Klasterizatsii parametrov bureniya. Tekhnologicheskoe razvitie nauki: tendentsii, problemy i perspektivy: sbornik statey po itogam Mezhdunarodnoy nauchno-prakticheskoy konferentsii* (Sterlitamak, June, 8). Ufa, Aeterna LLC, pp. 30-35. (In Russian).
21. Men'shikov, A. I. (2020). *Issledovanie rezhimov raboty sharoshechnykh dolot na osnove derev'ev resheniy. Informatsionnye tekhnologii kak osnova progressivnykh nauchnykh issledovaniy: sbornik trudov Mezhdunarodnoy nauchno-prakticheskoy konferentsii* (Perm, May, 25), Ufa, Aeterna LLC, pp. 78-83. (In Russian).
22. Men'shikov, A. I. (2020). *Prognozirovanie mekhanicheskoy skorosti bureniya massiva sharoshechnymi dolotami na osnove neyrosetevogo modelirovaniya. Informatsionnye tekhnologii kak osnova progressivnykh nauchnykh issledovaniy: sbornik trudov Mezhdunarodnoy nauchno-prakticheskoy konferentsii* (Perm, May, 25), Ufa, Aeterna LLC Publ., pp. 83-87. (In Russian).

Сведения об авторе

Катанов Юрий Евгеньевич, к. г.-м. н., доцент кафедры прикладной геофизики, ведущий научный сотрудник лаборатории технологий капитального ремонта скважин и интенсификации притока, Тюменский индустриальный университет, г. Тюмень, e-mail: katanov-juri@rambler.ru

Information about the author

Yury E. Katanov, Candidate of Geology and Mineralogy, Associate Professor at the Department of Applied Geophysics, Leading Researcher at Well Workover Technology and Production Stimulation Laboratory, Industrial University of Tyumen, e-mail: katanov-juri@rambler.ru