

Научная статья

УДК 622.276:004.896

doi: 10.17223/19988605/60/8

## Бинарная классификация скважин нефтегазовых промыслов с использованием глубоких нейронных сетей

Евсюткин Иван Викторович<sup>1</sup>, Марков Николай Григорьевич<sup>2</sup>

<sup>1,2</sup> *Томский политехнический университет, Томск, Россия*

<sup>1</sup> *iveI@tpu.ru*

<sup>2</sup> *markovng@tpu.ru*

**Аннотация.** Предложены модели глубоких искусственных нейронных сетей прямого распространения для решения задачи выбора скважин-кандидатов на проведение геолого-технических мероприятий на фонде скважин. Разработан адаптивный алгоритм предварительной обработки исходных данных. Приведены результаты исследования ряда моделей нейронных сетей на подготовленных с помощью этого алгоритма обучающей и тестовой выборках по фондам скважин двух месторождений. Выявлено подмножество моделей, позволяющих получить практически приемлемую точность классификации скважин-кандидатов.

**Ключевые слова:** скважина-кандидат для проведения геолого-технического мероприятия на фонде скважин; интеллектуальный анализ данных; бинарная классификация скважин; модели искусственных нейронных сетей

**Для цитирования:** Евсюткин И.В., Марков Н.Г. Бинарная классификация скважин нефтегазовых промыслов с использованием глубоких нейронных сетей // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 60. С. 73–83. doi: 10.17223/19988605/60/8

Original article

doi: 10.17223/19988605/60/8

## Binary classification for wells of oil and gas field with the use of deep neural networks

Ivan V. Evsyutkin<sup>1</sup>, Nikolay G. Markov<sup>2</sup>

<sup>1,2</sup> *Tomsk Polytechnic University, Tomsk, Russian Federation*

<sup>1</sup> *iveI@tpu.ru*

<sup>2</sup> *markovng@tpu.ru*

**Abstract.** The well-stock of the oil and gas extraction enterprise only on one field can include hundreds and even thousands of wells. The management of such well-stock demands considerable labor and time expenditure of qualified specialists of the extracting enterprise. They have to analyze large volumes of diverse geological and technological data. Part of the tasks solved by specialists at the management of a well-stock, including the management of the geological and technical arrangements (GTA) on the well-stock, are loosely formalized tasks. Intellectual methods of decision-making support are extremely necessary for experts for their solving. The analysis of research results in the field of intellectual methods for well-stock management showed that artificial neural networks (ANN) are generally applied. The first encouraging results are received using ANN, however, only partial tasks are solved and the accuracy of the received results is not high. It indicates the relevance of the development of new models and methods of intellectual data analysis for high-automated management of the well-stock.

The problem of candidates-wells selection for the GTA on a well-stock which can be simplified to the task of binary classification of production wells is considered in the article. It is offered to solve it with the use of the developed models of deep feed-forward ANN. Historical data from well-stocks of oil and gas-condensate fields of the Tomsk region are used as basic data for preparation of the training and evaluating selections for the ANN models.

The adaptive algorithm of preliminary processing of such data is developed, taking into account specifics of basic technological and geological data on the well-stocks. The training and evaluating selections for the offered ANN models are created with the algorithm's help. The following architectural characteristics and hyper-parameters varied at the research of efficiency of the ANN models: the set of the input parameters, the number of the hidden layers, an activation function, a training speed, an optimizing training algorithm, etc.

Research results of the set of the offered ANN models on the prepared training and evaluating selections on the production well-stocks of these two fields showed that there is a subset of the ANN models allowing to receive high (98,5 % and above) the precision of classification of candidates-wells for GTA, acceptable for practical use on oil and gas fields at GTA management. These results also emphasize adequacy of the revealed subset of the ANN models of the complexity of the solved classification task without dependence on the prevailing fluid in the got raw oil and gas products on a field and the geological structure of a field. It is shown that the ANN models, trained on the joint selection by data from well-stocks of production wells of two fields, generally give more a good result on the accuracy of classification of wells, than in the case of their training on data from one field.

**Keywords:** a candidate-well for carrying out geological and technical arrangements on a well-stock; intellectual analysis of data; the binary classification of wells; the models of artificial neural networks

**For citation:** Evsyutkin, I.V., Markov, N.G. (2022) Binary classification for wells of oil and gas field with the use of deep neural networks. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 60. pp. 73–83. doi: 10.17223/19988605/60/8

Фонд скважин нефтегазодобывающего предприятия только на одном месторождении может включать сотни и даже тысячи скважин. Столь крупный объект управления требует серьезного внимания со стороны субъектов управления – служб предприятия – с целью формирования и оказания различных управляющих воздействий на скважины фонда и продуктивные пласты для поддержания добычи углеводородного сырья (УВС) на проектных уровнях. Управление фондом скважин промысла требует значительных трудовых и временных затрат квалифицированных специалистов добывающего предприятия, которые должны анализировать большие объемы разнородных геологических и технологических данных, получаемых при мониторинге продуктивных пластов месторождения и скважин фонда [1]. Однако существующие сегодня у большинства добывающих предприятий производственные информационные системы (ИС) являются чаще всего узкоспециализированными, и в них не реализованы интеллектуальные методы поддержки принятия решений специалистами при управлении фондом скважин [2]. Часть решаемых специалистами задач при управлении фондом скважин, в том числе при управлении геолого-техническими мероприятиями (ГТМ), – это слабоформализуемые задачи [3]. При их решении специалистам необходимы интеллектуальные методы поддержки принятия решений. По мнению ряда исследователей [4, 5], сегодня методы и алгоритмы интеллектуального анализа данных (ИАД) являются альтернативными многим традиционным методам и алгоритмам, применяемым в нефтегазовой отрасли.

В работе [6] приведены результаты исследования эффективности искусственных нейронных сетей (ИНС) для прогноза параметров работы скважин после проведения гидроразрыва пласта. На нескольких нефтяных месторождениях ООО «Газпромнефть-Ноябрьскнефтегаз» проведен достаточно большой объем работ по гидроразрыву пластов. Результаты прогноза с помощью модели ИНС в виде персептрона на этих данных оказались невысокими. В [7] модели ИНС использовались для анализа операций по выравниванию профиля приемистости для нагнетательных скважин, а также для оценки влияния геолого-геофизических характеристик продуктивных пластов и технологических параметров скважин на результат в виде прироста уровня добычи нефти вследствие уменьшения обводненности УВС. При этом точность прогноза при использовании моделей ИНС оказалась сопоставимой с таковой у метода гидродинамического моделирования, но скорость выполнения моделей ИНС гораздо выше. При решении той же задачи выравнивания профиля приемистости предварительно был использован метод кластеризации для группировки исходных данных на нефтяных месторождениях Поволжья [8]. Это позволило несколько уменьшить ошибку в предсказаниях характеристик добывающих скважин. К сожалению, все полученные результаты прогноза далеки от результатов, представляющих практический интерес. Из имеющихся обзоров следует, что интересные результаты при использовании ИНС в нефтегазовой

отрасли получены в транснациональной компании Schlumberge, однако особенности применяемых в компании моделей ИНС в литературе практически не описаны. Подводя итог, можно сказать, что в нефтегазовой отрасли получены первые обнадеживающие результаты при решении частных задач с помощью методов ИАД. Однако точность этих результатов невысока, а решение ряда практически важных задач управления фондом скважин с помощью таких методов даже не рассматривалось. Все это указывает на актуальность разработки новых моделей, методов и алгоритмов ИАД и их программной реализации для высокоавтоматизированного управления фондом скважин в условиях постоянно обновляющихся на промыслах данных.

Одной из ключевых слабоформализуемых задач при управлении фондом скважин является задача выбора скважин-кандидатов для ГТМ, сводящаяся к задаче бинарной классификации всех добывающих скважин фонда. В работах, посвященных решению этой задачи, получены первые результаты применения моделей ИНС на нефтяных (основной добываемый флюид – нефть) месторождениях [9]. В действительности же каждое разрабатываемое месторождение имеет весьма сложное геологическое строение (разное число продуктивных пластов и пропластков, разная геометрия резервуаров (пластов) и т.п.) и разный состав и свойства добываемого УВС. В связи с этим актуальным является проведение исследований, показывающих возможности предлагаемых моделей ИНС для ИАД по фондам скважин месторождений с другим геологическим строением и с другим типом сырья (основной добываемый флюид – газ или газовый конденсат), чем у месторождения в [9].

В статье приводятся результаты решения такой задачи с использованием разработанных моделей глубоких ИНС прямого распространения. Предварительная подготовка данных с целью формирования обучающей и тестовой выборок для ИНС ведется с помощью разработанного адаптивного алгоритма. Показано, что обеспечиваемая с помощью разработанного набора моделей ИНС точность решения задачи бинарной классификации добывающих скважин для различных по геологическому строению и по приоритетному типу флюида месторождений удовлетворяет практически важным требованиям специалистов добывающих предприятий.

## **1. Задача выбора скважин-кандидатов для ГТМ**

Известно, что важной задачей при управлении фондом скважин является задача выбора скважин-кандидатов для ГТМ, сводящаяся к задаче бинарной классификации всех добывающих скважин фонда. Предлагается решать эту задачу с помощью ИНС. В первый класс должны попасть все скважины, которые планируются к проведению определенного типа (вида) ГТМ. Это класс «скважин-кандидатов для ГТМ». Основанный на опыте и интуиции специалистов и проведенный ранее на фонде ручной выбор скважин-кандидатов специалистами-геологами будет считаться эталоном при формировании обучающей выборки для ИНС, если после проведенного ГТМ произошло существенное увеличение дебита скважины. Во второй класс скважин отнесем все остальные скважины фонда (класс «скважин не для ГТМ»), на которых в момент времени, когда принималось решение, по той или иной причине ГТМ проводить не следовало. Например, это высокодебитные скважины, дающие основной объем добычи УВС на месторождении.

Большинство исследователей при выборе модели ИНС руководствуются правилом соответствия сложности задачи и функциональной мощности выбираемого инструмента (модели ИНС) для ее решения: не следует использовать сложную модель ИНС при решении простой задачи. Были проанализированы различные типы (классы) ИНС, и в качестве наиболее перспективного типа для решения поставленной задачи классификации скважин выбраны глубокие ИНС прямого распространения FFNN [10]. Такой тип моделей ИНС является компромиссом между невысокой сложностью моделей ИНС, но достаточными их функциональными возможностями при решении стоящей задачи. По сути, делается допущение, что возможностей таких нейронных сетей будет достаточно при решении задачи выбора скважин-кандидатов для ГТМ. Однако необходимо исследовать эффективность FFNN для ответа на вопрос, может ли этот тип (класс) ИНС дать точность при решении задачи классификации скважин, приемлемую с практической точки зрения.

Подготовка обучающей и тестовой выборки для моделей ИНС ведется на основе исторических данных из баз данных (БД) предприятия по разрабатываемым месторождениям. Такие данные обычно получены путем измерения значений ряда технологических и геологических параметров фонда скважин месторождения, для которого необходимо решить задачу классификации и сформировать список скважин-кандидатов для ГТМ. Среди таких параметров наиболее важными являются геологические и технологические параметры каждого комплекса «скважина–продуктивный пласт»: дебит нефти  $Q_n$ , т; дебит газа  $Q_g$ , тыс. м<sup>3</sup>; дебит жидкости  $Q_j$ , м<sup>3</sup>; дебит воды  $Q_v$ , т; давление забойное  $P_{заб}$ , атм; давление буферное  $P_{буф}$ , атм; давление затрубное  $P_{зат}$ , атм; температура на устье скважины  $T_y$ , °С; давление пластовое  $P_{пл}$ , атм; пластовая температура  $T_{пл}$ , °С; обводненность  $W$ , %; эффективная мощность пласта  $H$ , м; газовый фактор  $B$ ; диаметр штуцера  $D$ , мм. Скважины, признанные специалистами геологических служб предприятия кандидатами на ГТМ и на которых были проведены ГТМ с увеличением после мероприятия уровня добычи УВС, считались нами эталонными и имели при подготовке выборки соответствующие метки.

При исследовании эффективности модели ИНС важно понять, какое сочетание геологических и (или) технологических параметров содержит наиболее существенные признаки, которые модель ИНС может извлечь из них для решения поставленной задачи классификации. Формировались следующие обучающие и тестовые наборы параметров, подаваемые затем на вход ИНС:

Набор № 1 – все чаще всего измеряемые на промыслах параметры скважин и продуктивных пластов:  $Q_n$ ;  $Q_g$ ;  $Q_j$ ;  $Q_v$ ;  $P_{заб}$ ;  $P_{буф}$ ;  $P_{зат}$ ;  $T_y$ ;  $D$ ;  $P_{пл}$ ;  $T_{пл}$ ;  $W$ ;  $H$ ;  $B$ .

Набор № 2 – все технологические параметры скважин:  $Q_n$ ;  $Q_g$ ;  $Q_j$ ;  $Q_v$ ;  $P_{заб}$ ;  $P_{буф}$ ;  $P_{зат}$ ;  $T_y$ ;  $B$ .

Набор № 3 – параметры дебитов скважин:  $Q_n$ ;  $Q_g$ ;  $Q_j$ ;  $Q_v$ .

Набор № 4 – технологические параметры скважин (без дебитов):  $P_{заб}$ ;  $P_{буф}$ ;  $P_{зат}$ ;  $T_y$ ;  $B$ .

Набор № 5 – параметры дебитов скважин и параметры продуктивных пластов:  $Q_n$ ;  $Q_g$ ;  $Q_j$ ;  $Q_v$ ;  $P_{пл}$ ;  $T_{пл}$ .

Точность классификации добывающих скважин фонда с помощью модели ИНС из класса FFNN должна оцениваться для обучающей выборки и для тестовой выборки отдельно путем вычисления специальной метрики [11]. Несмотря на интуитивную понятность наиболее известной из таких метрик – метрики *Accuracy* (точность), при ее применении требуется сбалансированность выборки для каждого из классов при бинарной классификации: оба класса скважин должны быть представлены выборками приблизительно равного объема. В реальности же так не всегда получается, в первую очередь из-за того, что ГТМ на добывающей скважине проводятся гораздо реже, чем скважина работает без ГТМ. Поэтому использовалась метрика *Precision*, часто применяемая при решении задач классификации объектов в случаях, когда выборки по классам объектов значительно отличаются по объему. Эта метрика позволяет вычислять долю скважин, в реальности принадлежащих данному классу («скважина-кандидат для ГТМ» или «скважина, на которой ГТМ проводить не следует»), относительно всех скважин, которые модель ИНС отнесла к этому классу:

$$Precision = \frac{TP}{TP + FP} \times 100 \%, \quad (1)$$

где  $TP$  – истинно-положительное решение;  $FP$  – ложно-положительное решение.

Отметим, что далее с помощью выражения (1) получены оценки точности классификации скважин для класса «скважина-кандидат для ГТМ».

## 2. Формирование моделей ИНС для исследования

Выбранный класс моделей FFNN имеет большое число варьируемых характеристик архитектуры и гиперпараметров, что позволяет создавать и обучать в рамках этого класса различные модели ИНС. Каждая из таких моделей ИНС описывается парой множеств:

$$\langle Architecture, Hyperparameters \rangle, \quad (2)$$

где *Architecture* – множество характеристик архитектуры ИНС (число подаваемых на вход технологических и геологических параметров, число скрытых слоев, число нейронов в скрытых слоях, функ-

ция активации и т.д.); *Hyperparameters* – множество гиперпараметров ИНС при обучении (скорость обучения, алгоритм обучения, размер мини-выборки, число эпох).

Среди наиболее часто варьируемых характеристик – число скрытых слоев FFNN, число значений геологических и технологических параметров во входном слое ИНС и функция активации нейронов. Рациональный выбор характеристик и параметров в выражении (2) должен обеспечивать формирование модели (моделей) ИНС и ее (их) обучение с целью решения с требуемой точностью задачи классификации. Число сочетаний различных параметров и характеристик модели ИНС может быть весьма велико, поэтому очень важно ограничить задачу поиска только теми из них и, соответственно, остановиться на тех моделях ИНС, которые с высокой вероятностью дают гарантированный результат для практического использования при решении задачи бинарной классификации для фондов скважин на месторождениях с различными геологическим строением и преобладающим типом флюида (нефть, газ или газовый конденсат).

Число скрытых слоев в моделях ИНС при исследовании выбиралось равным 2, 3, 4, ..., 9, 10, 15, 20. Формировались модели ИНС с различным числом входов. В исследованиях осуществлялась поочередная подача на вход ИНС каждого из пяти сформированных наборов геологических и технологических параметров, поэтому число входов модели ИНС определялось числом параметров в наборе. Для всех таких моделей ИНС исследовался эффект от применения различных функций активации (Sigmoid, TanH, ReLU) и алгоритмов оптимизации при обучении ИНС (SGD, Adam, AdaGrad, AdaDelta). Изменялась также скорость обучения: 0,2; 0,1; 0,05; 0,01; 0,005; 0,001. Обобщенная модель включает в себя входной слой, куда подаются значения параметров из того или иного обучающего или тестового набора из пяти перечисленных для исследуемой скважины, совокупность скрытых слоев с конкретной функцией активации и выходной слой (Softmax). Выходной слой содержит два нейрона, позволяющих сделать вывод, является ли скважина кандидатом на ГТМ. Для предотвращения переобучения перед выходным слоем в модели ИНС использовался слой Dropout [12].

Программная реализация моделей ИНС осуществлялась с использованием нейросетевой библиотеки CNTK компании Microsoft [13].

### 3. Предварительная обработка данных

Решение поставленной задачи классификации осуществлялось с использованием архивов данных в виде БД по двум эксплуатируемым месторождениям: нефтяному месторождению (Месторождение 1) и газоконденсатному месторождению (Месторождение 2), находящимся в Томской области. Фонд скважин Месторождения 1 имеет 142 добывающих скважины, на которых было проведено 436 различных ГТМ за 6 лет. Фонд скважин Месторождения 2 имеет 60 добывающих скважин, на которых было проведено 313 различных ГТМ за 5 лет. В БД за несколько лет эксплуатации месторождений накоплен значительный объем данных по параметрам каждого комплекса «скважина–продуктивный пласт» и имеется информация о проведенных ГТМ.

Исходные данные для последующего ИАД с целью выработки управляющих воздействий на фонд скважин могут поступать (собираться) в БД из различных ИС и (или) АСУ ТП предприятия. Зачастую такая БД содержит ошибочную, пропущенную и искаженную информацию о параметрах скважин фонда и продуктивных пластов эксплуатируемого месторождения. Это означает, что в соответствии с концепцией Big Data необходима предварительная обработка (актуализация) исходных данных с таких месторождений. При этом следует использовать различные подходы, методы и алгоритмы обработки постоянно растущих объемов данных. Анализ показал, что данные из БД по Месторождению 1 и Месторождению 2 не являются исключением и также подлежат предварительной обработке (подготовке), прежде чем будут использованы при обучении и тестировании моделей ИНС. Формирование обучающей и тестовой выборки проводилось с учетом особенностей геологических и технологических данных по разработанному и программно реализованному адаптивному алгоритму. Алгоритм позволяет также отбраковывать ошибочные данные, анализировать корректность данных, определять характер пропусков в данных и с учетом этого адаптивно восстанавливать их с требуемой

точностью. Он отличается от известных методов и алгоритмов наличием решающих правил, позволяющих адаптировать процесс вычисления при восстановлении данных к видам пропусков и ошибок в исходных данных. Рассмотрим алгоритм более подробно.

**Начало.**

**Шаг 1.** С помощью SQL-запросов к БД выбранного месторождения происходит извлечение (сбор) данных с целью их последующей предварительной обработки.

**Шаг 2.** Ведется выбраковка из собранных данных явно ошибочных значений параметров (обычно допущены при ручном вводе в БД), далее в анализе они не используются; оставшиеся данные сводятся в две таблицы, причем первая из них содержит паспортные данные скважин и значения технологических и геологических параметров в виде временных рядов, а вторая таблица включает данные, связанные с проведенными на скважинах и в продуктивных слоях различными ГТМ.

**Шаг 3.** Ведется анализ данных на корректность, при этом с учетом диапазонов значений параметров для конкретного месторождения определяются ошибочные или пропущенные данные, а также аномально большие отсчеты параметров из-за сбоев измерительной аппаратуры; поиск таких данных осуществляется во многих случаях путем сравнения с соседними во временном ряду значениями параметра. Найденные значения удаляются, но вместо них на Шаге 4 вводятся новые.

**Шаг 4.** Адаптивное восстановление с использованием решающих правил удаленных или пропущенных данных и интерполяция значений некоторых параметров по соседним имеющимся значениям для получения вектора значений всех параметров на определенный момент времени.

**Шаг 5.** Формирование массива обучающих примеров; проводится с учетом периода выполнения ГТМ на добывающей скважине или времени ее простоя по другим причинам, поскольку в вектор значений параметров на вход ИНС по каждой скважине не должны попасть значения одновременно до и после проведения тех или иных ГТМ или остановок скважины.

**Шаг 6.** Создание дополнительных обучающих примеров из имеющегося массива обучающих примеров с помощью метода аугментации – добавления к значениям ряда технологических и геологических параметров в Наборах № 1–5 случайного 5%-ного шума [14].

**Шаг 7.** Нормализация всех подготовленных данных, поскольку значения параметров скважин и продуктивных пластов измеряются в разных диапазонах, а каждый параметр должен оказывать влияние на конечный результат в исследованиях моделей ИНС.

**Шаг 8.** Формируются обучающая и тестовая выборки: все полученные обучающие примеры относятся к одной из них, при этом тестовая выборка составляет 20% от общего числа примеров, а обучающая – 80%.

**Конец.**

*Замечание 1.* Разработанные решающие правила на Шаге 4 учитывают все возможные ситуации (единичный пропуск, совокупность пропущенных или удаленных значений параметра, размер интервала пропуска и т.п.) и позволяют выбрать и запустить один из реализованных методов интерполяции, дающий в конкретном случае наиболее высокую точность для восстанавливаемых данных. Реализованы методы линейной интерполяции, локальной кусочно-квадратичной интерполяции и одномерный кубический сплайн [15]. Разработка решающих правил проведена с использованием результатов исследований точности каждого из этих методов на данных по фонду скважин Месторождения 1 [16].

*Замечание 2.* Процедура получения дополнительных обучающих примеров на Шаге 6 алгоритма с использованием метода аугментации данных весьма важна. Во-первых, возникает проблема, если фонд скважин месторождения имеет относительно небольшое количество добывающих скважин (особенно если это месторождение эксплуатируется не так давно), а обучающая выборка для получения высокой точности классификации скважин с помощью ИНС должна иметь значительный объем. Во-вторых, в любом фонде тех скважин, которые подвергались ГТМ в выбранный период времени, обычно меньше, поэтому существует проблема несбалансированности выборок для двух классов скважин. Эти две проблемы и решаются с помощью указанного метода аугментации имеющихся данных.

#### 4. Результаты исследования моделей ИНС

С использованием приведенного алгоритма осуществлена предварительная обработка данных из БД фондов скважин Месторождения 1 и Месторождения 2. Для первого из них число обучающих примеров, включая полученные с помощью метода аугментации, составило 2 016 пар «признаки–метки» для каждого из пяти указанных выше наборов параметров, что позволило сформировать обучающие и тестовые выборки для моделей ИНС в случае каждого из пяти наборов параметров. К сожалению, не для всех вышеперечисленных параметров фонда скважин Месторождения 2 имеется достаточный объем данных, поэтому для его комплексов «скважина–продуктивный пласт» в исследованиях использовался урезанный набор из таких параметров. Удалось провести предварительную обработку данных, подготовить 3 371 обучающий пример и сформировать обучающую и тестовую выборки только в случае параметров из Набора № 4.

Для повышения точности классификации скважин при проведении исследований моделей ИНС использовался один из методов кросс-валидации – метод *K* блоков [17], причем с учетом того, что отношение объемов обучающей и тестовой выборок выбрано 4:1, число блоков *K* равно 5. При формировании первого блока в тестовую выборку попадает каждый пятый обучающий пример из общего объема обучающей и тестовой выборок, а при построении очередного блока происходит то же самое, но все номера обучающих примеров изменяются на единицу и т.д. Точность классификации при обучении и тестировании модели ИНС оценивается в случае каждого блока в виде значений метрики *Precision*, вычисляемой по формуле (1) для практически важного класса «скважина-кандидат для ГТМ». Затем вычисляется усредненная метрика *Precision* с учетом результатов по каждому из пяти блоков.

Первые результаты исследований эффективности каждой модели ИНС при решении поставленной задачи классификации были получены для Месторождения 1 и приведены в [9]. Анализ этих результатов позволил сделать вывод, что основным фактором, влияющим на точность классификации, является различие моделей ИНС, в первую очередь по числу скрытых слоев. Точность также в значительной мере определяется перечнем параметров, подаваемых на вход ИНС. При этом наилучшая точность достигнута для Наборов № 1 и № 2, а наихудшая – для Набора № 3.

В табл. 1 показаны результаты экспериментов с моделями ИНС с изменяемым числом скрытых слоев при решении задачи классификации скважин в случае фонда Месторождения 2 для тестовой выборки. Аналогичная таблица с результатами для точности классификации получена в случае обучающей выборки. Значения точности классификации в ней на 1–1,5% выше, что указывает на правильно полученные в процессе обучения архитектуру и гиперпараметры ИНС. При этом взвешенные средние абсолютные погрешности составляют соответственно 0,3 и 0,7% для обучающей и тестовой выборок.

Таблица 1

Результаты классификации скважин-кандидатов для ГТМ на фонде Месторождения 2

Алгоритм обучения	Функция активации	Точность классификации добывающих скважин для различного числа скрытых слоев в модели ИНС, %											
		2	3	4	5	6	7	8	9	10	15	20	
SGD	Sigmoid	79,111	80,277	81,630	81,333	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185
	TanH	94,519	98,222	97,926	<b>98,815</b>	97,778	91,926	<b>98,512</b>	95,259	95,111	89,630	97,926	
	ReLU	93,333	97,926	98,815	<b>99,259</b>	<b>98,667</b>	94,815	<b>98,667</b>	<b>98,667</b>	97,630	97,333	95,111	
AdaGrad	Sigmoid	77,037	77,630	78,370	74,963	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185
	TanH	92,296	<b>98,519</b>	<b>99,259</b>	<b>98,815</b>	<b>99,111</b>	<b>98,963</b>	96,889	<b>99,259</b>	<b>98,815</b>	<b>98,963</b>	96,148	
	ReLU	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185
AdaDelta	Sigmoid	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185	53,185
	TanH	74,963	73,481	73,630	73,185	74,667	74,222	74,667	74,815	75,407	75,259	79,407	
	ReLU	73,481	74,593	76,296	77,630	77,778	77,630	78,519	80,000	80,296	80,889	81,630	
Adam	Sigmoid	89,778	91,704	95,407	94,222	92,000	93,481	85,481	87,852	83,259	53,185	53,185	
	TanH	98,370	98,370	<b>98,666</b>	<b>98,666</b>	<b>98,666</b>	95,111	96,296	93,185	85,333	46,815	46,815	
	ReLU	97,333	90,074	98,222	<b>98,519</b>	<b>98,519</b>	<b>98,963</b>	97,778	98,074	97,037	54,222	53,185	

Из табл. 1 следует, что для Набора № 4 параметров для некоторого числа скрытых слоев и функций активации TanH и ReLU получены весьма высокие результаты (98,5% и выше, что приемлемо на практике; такие значения выделены жирным шрифтом) при использовании алгоритмов обучения SGD, AdaGrad и Adam. Из сравнения этих результатов с результатами для фонда Месторождения 1 в случае набора параметров № 4, изложенными в [9], следует, что модели ИНС, которые показали невысокие результаты на данных по фонду Месторождения 1, чаще всего дают невысокие результаты и для фонда Месторождения 2.

Был проведен ряд дополнительных исследований. Во-первых, модели ИНС, обученные на данных по фонду скважин Месторождения 1, были исследованы на тестовой выборке Месторождения 2. При этом наряду с обученными моделями ИНС для фонда Месторождения 1, которые дали самую высокую точность на тестовой выборке этого фонда, исследовались модели ИНС, которые дали на этой же тестовой выборке невысокую точность. Наоборот, обученные модели ИНС для фонда Месторождения 2 были исследованы на тестовой выборке Месторождения 1. Во-вторых, было осуществлено слияние обучающих выборок с фондов скважин этих месторождений. Несмотря на то, что для фонда Месторождения 2 имеется большая обучающая выборка, чем для фонда Месторождения 1, объединенная выборка была использована для обучения моделей ИНС и для Месторождения 2. Результаты этих исследований на соответствующей тестовой выборке каждого из месторождений представлены в табл. 2. Жирным шрифтом выделены лучшие результаты, полученные в случае обучения ИНС на объединенной выборке.

Таблица 2

Результаты классификации скважин-кандидатов для ГТМ на фондах Месторождений 1 и 2 с использованием данных и моделей ИНС другого месторождения

Алгоритм обучения	Функция активации	Число скрытых слоев	Точность классификации скважин, %					
			Месторождение 1			Месторождение 2		
			Обучение на своих данных	Обучение на объединенной выборке	Использованы обученные модели Мест. 2	Обучение на своих данных	Обучение на объединенной выборке	Использованы обученные модели Мест. 1
SGD	Sigmoid	3	76,877	79,558	61,086	80,277	81,630	61,187
	TanH	10	94,836	<b>99,334</b>	54,102	95,111	<b>98,963</b>	58,293
	ReLU	6	95,572	<b>99,533</b>	72,931	98,667	<b>99,259</b>	62,857
AdaGrad	Sigmoid	2	76,753	75,691	61,197	77,037	77,481	59,740
	TanH	6	95,203	<b>99,933</b>	52,239	99,111	98,963	63,636
	ReLU	4	58,426	60,773	60,754	53,185	53,185	53,210
AdaDelta	Sigmoid	4	58,426	60,773	60,754	53,185	53,185	53,210
	TanH	8	71,834	71,242	42,749	74,667	74,667	67,199
	ReLU	10	66,053	77,901	56,785	80,296	80,296	53,210
Adam	Sigmoid	6	87,824	95,301	61,552	92,000	96,593	55,547
	TanH	3	97,418	<b>99,800</b>	58,714	98,370	<b>98,667</b>	62,894
	ReLU	4	97,173	<b>99,448</b>	63,282	98,222	<b>98,963</b>	62,894

Из табл. 2 видно, что модели ИНС, обученные на объединенной выборке, полученной для фондов месторождений даже с разным составом добываемого УВС и с различным геологическим строением месторождений, дают более высокий результат по точности классификации скважин, чем в случае их обучения на данных с одного месторождения. Однако если с использованием обучающей выборки месторождения уже были получены высокие результаты по точности классификации, то выигрыша от объединения выборок может не быть. Модели ИНС, построенные и обученные на данных одного фонда скважин, при тестировании их на данных другого фонда дают более низкую точность классификации скважин (см. третий столбец для каждого из месторождений), чем при тестировании их на данных с собственного фонда. По-видимому, извлекаемые ИНС признаки, которыми обладают скважины-кандидаты для ГТМ конкретного фонда, являются довольно индивидуальными. Это не позволяет применять обученные на одном фонде модели ИНС при решении задач классификации скважин на фондах других месторождений без дообучения их на данных этих фондов скважин.

Приведенные результаты позволяют сделать вывод, что на множестве исследованных моделей ИНС прямого распространения существует подмножество моделей, дающих высокую (98,5% и выше) с практической точки зрения точность классификации добывающих скважин вне зависимости от преобладающего типа флюида в добываемом УВС (в случае Месторождения 1 – нефть, а в случае Месторождения 2 – газ) и вне зависимости от геологического строения этих месторождений. Отметим, что, по мнению специалистов-практиков с нефтегазовых промыслов, такая точность классификации скважин полностью удовлетворяет их требованиям.

### **Заключение**

Рассмотрено применение глубоких ИНС прямого распространения для решения задачи выбора скважин-кандидатов для ГТМ на фонде добывающих скважин месторождения. С учетом специфики исходных технологических и геологических данных по фонду скважин и принципов концепции Big Data разработан адаптивный алгоритм предварительной обработки данных с целью формирования обучающей и тестовой выборок для таких моделей ИНС.

Результаты исследования множества предложенных моделей ИНС прямого распространения на подготовленных с помощью адаптивного алгоритма обучающей и тестовой выборках по фондам добывающих скважин нефтяного и газоконденсатного месторождений показали, что найдено подмножество моделей ИНС, позволяющих получить высокую (98,5% и выше) точность классификации скважин-кандидатов для ГТМ. Такая точность приемлема для практического использования моделей ИНС на промыслах при управлении ГТМ. Эти результаты также подчеркивают адекватность выявленного подмножества моделей ИНС сложности решаемой задачи классификации независимо от преобладающего флюида в добываемом УВС и от геологического строения месторождений.

Модели ИНС, обученные на объединенной выборке по данным из фондов добывающих скважин двух месторождений, дают в общем случае более высокий результат по точности классификации скважин, чем в случае их обучения на данных одного фонда. Все приведенные результаты в итоге позволяют считать, что предлагаемые наборы моделей ИНС будут обеспечивать точность классификации скважин, приемлемую для использования с практической точки зрения на разных по геологическому строению и составу флюидов месторождениях.

### **Список источников**

1. Шадрин А.В., Крец В.Г. Основы нефтегазового дела. М. : Нац. Открытый Ун-т «ИНТУИТ», 2016. 214 с.
2. Марков Н.Г. Информационно-управляющие системы для газодобывающего производства. Томск : Изд-во Том. политех. ун-та, 2016. 261 с.
3. Евсюткин И.В., Марков Н.Г. Глубокие искусственные нейронные сети для прогноза значений дебитов добывающих скважин // Известия Томского политехнического университета. Инжиниринг георесурсов. 2020. Т. 331, № 11. С. 88–95.
4. Taha A., Amani M. Introduction to Smart Oil and Gas Wells: Drilling, Completion and Monitoring Solutions // Int J Petrochem Res. 2019. V. 3, № 1. P. 249–254.
5. Redouane S., Suresh R., Subashini S., Saeed Al N. Smart solutions in the oil and gas industry: a review // Journal of Clean Energy Technologies. 2019. V. 7, № 5. P. 72–76.
6. Дмитриевский А.Н. Цифровизация и интеллектуализация нефтегазовых месторождений // Автоматизация и информационные технологии в нефтегазовой области. 2016. Т. 2, № 24. С. 13–19.
7. Келлер Ю.А. Разработка искусственных нейронных сетей для предсказания технологической эффективности от выравнивания профиля приемистости // Известия Томского политехнического университета. Информационные технологии. 2014. Т. 325, № 5. С. 60–65.
8. Келлер Ю.А. Применение кластеризации данных на основе самоорганизующихся карт Кохонена при подборе скважин-кандидатов для методов увеличения нефтеотдачи // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2014. Т. 28, № 3. С. 32–37.
9. Евсюткин И.В., Марков Н.Г. Управление геолого-техническими мероприятиями на месторождениях нефти и газа с использованием искусственных нейронных сетей // Доклады ТУСУР. Управление, вычислительная техника и информатика. 2020. Т. 23. № 1. С. 62–69.
10. Schmidhuber J. Deep Learning in Neural Networks: an Overview // Neural Networks. 2015. № 61. P. 85–117.
11. Juba B., Le H.S. Precision-Recall versus Accuracy and the Role of Large Data Sets // Proc. of the AAAI Conference on Artificial Intelligence. 2019. V. 33, № 01. P. 4039–4048.

12. Ma Z., Sattar A., Zhou J., Chen Q., Su K. Dropout with Tabu Strategy for Regularizing Deep Neural Networks // *The Computer Journal*. 2020. V. 63, № 7. P. 1031–1038.
13. Hatcher W.G., Yu W.A. Survey of Deep Learning: Platforms, Applications and Emerging Research Trends // *IEEE Access*. 2018. V. 6. P. 24411–24432.
14. Khandakar M., Louis J. Times-series data augmentation and deep learning for construction equipment activity recognition // *Advanced Engineering Informatics*. 2019. V. 42. Art. 100944. 12 p.
15. Шевченко А.С. Численные методы. М. : ИНФРА-М, 2021. 381 с.
16. Evsyutkin I.V., Markov N.G. The intellectual analysis of geological and technological data during the management of an oil field's well-stock // *Journal of Physics: Conference Series*. 2020. V. 1661. Art. 012033. 7 p.
17. Ahmed F.Y.H., Ali Y.H., Shamsuddin S.M. Using K-Fold Cross Validation Proposed Models for Spikeprop Learning Enhancements // *International Journal of Engineering & Technology*. 2018. V. 7, № 4. P. 145–151.

### References

1. Shadrina, A.V. & Krets, V.G. (2016) *Osnovy neftegazovogo dela* [Fundamentals of Oil and Gas Engineering]. Moscow: National Open University of INTUIT.
2. Markov, N.G. (2016) *Informatsionno-upravlyayushchie sistemy dlya gazodobyvayushchego proizvodstva* [Information Control Systems for Gas Extraction Industry]. Tomsk: Tomsk Polytechnic University.
3. Evsyutkin, I.V. & Markov, N.G. (2020) Deep artificial neural networks for forecasting debit values for production wells. *Izvestiya Tomskogo politekhnicheskogo universiteta. Inzhiniring geosurosov – Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*. 331(11). pp. 88–95. DOI: 10.18799/24131830/2020/11/2888
4. Taha, A. & Amani, M. (2019) Introduction to Smart Oil and Gas Wells: Drilling, Completion and Monitoring Solutions. *International Journal of Petrochemistry and Research*. 3(1). pp. 249–254. DOI: 10.18689/ijpr-1000143
5. Redouane, S., Suresh, R., Subashini, S. & Saeed, Al.N. (2019) Smart solutions in the oil and gas industry: A review. *Journal of Clean Energy Technologies*. 7(5). pp. 72–76. DOI: 10.18178/JOCET.2019.7.5.512
6. Dmitriyevsky, A.N. (2016) Tsifrovizatsiya i intellektualizatsiya neftegazovykh mestorozhdeniy [Digitalization and intellectualization of oil and gas fields]. *Avtomatizatsiya i informatsionnye tekhnologii v neftegazovoy oblasti*. 2(24). pp. 13–19.
7. Keller, Yu.A. (2014) Razrabotka iskusstvennykh neyronnykh setey dlya predskazaniya tekhnologicheskoy effektivnosti ot vyravni-vaniya profilya priemistosti [Design of artificial neural networks for predicting the technological efficiency of improving water injection profile]. *Izvestiya Tomskogo politekhnicheskogo universiteta. Informatsionnye tekhnologii*. 325(5). pp. 60–65.
8. Keller, Yu.A. (2014) The application of data clustering on the basis of kohonen self-organizing maps in the process of selecting candidate wells for enhanced oil recovery methods. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(28). pp. 32–37.
9. Evsyutkin, I.V. & Markov, N.G. (2020) Management of geological and technical arrangements on oil-and-gas fields with the use of artificial neural networks. *Doklady TUSUR. Upravlenie, vychislitel'naya tekhnika i informatika – Proceedings of TUSUR University. Control, Computer Machines and Informatics*. 23(1). pp. 62–69.
10. Schmidhuber, J. (2015) Deep Learning in Neural Networks: An Overview. *Neural Networks*. 61. pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003
11. Juba, B. & Le, H.S. (2019) Precision-Recall versus Accuracy and the Role of Large Data Sets. *Proceedings of the AAAI Conference on Artificial Intelligence*. 33(01). pp. 4039–4048. DOI: 10.1609/aaai.v33i01.33014039
12. Ma, Z., Sattar, A., Zhou, J., Chen, Q. & Su, K. (2020) Dropout with Tabu Strategy for Regularizing Deep Neural Networks. *The Computer Journal*. 63(7). pp. 1031–1038. DOI: 10.48550/arXiv.1808.09907
13. Hatcher, W.G. & Yu, W. (2018) A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends. *IEEE Access*. 6. pp. 24411–24432.
14. Khandakar, M. & Louis, J. (2019) Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics*. 42(100944). DOI: 10.1016/j.aei.2019.100944
15. Shevchenko, A.S. (2021) *Chislennyye metody* [Numerical Techniques]. Moscow: INFRA-M.
16. Evsyutkin, I.V. & Markov, N.G. (2020) The intellectual analysis of geological and technological data during the management of an oil field's well-stock. *Journal of Physics: Conference Series*. 1661(012033). DOI: 10.1088/1742-6596/1661/1/012033
17. Ahmed, F.Y.H., Ali, Y.H., & Shamsuddin, S.M. (2018) Using K-Fold Cross Validation Proposed Models for Spikeprop Learning Enhancements. *International Journal of Engineering & Technology*. 7(4). pp. 145–151. DOI: 10.14419/ijet.v7i4.11.20790

### Информация об авторах:

**Евсюткин Иван Викторович** – кандидат технических наук, доцент Отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета (Томск, Россия). E-mail: ive1@tpu.ru

**Марков Николай Григорьевич** – профессор, доктор технических наук, профессор Отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета (Томск, Россия). E-mail: markovng@tpu.ru

***Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.***

***Information about the authors:***

**Evsyutkin Ivan V.** (Candidate of Technical Sciences, Assistant Professor, National Research Tomsk Polytechnic University, Tomsk, Russian Federation). E-mail: ive1@tpu.ru

**Markov Nikolay G.** (Doctor of Technical Sciences, Professor, National Research Tomsk Polytechnic University, Tomsk, Russian Federation). E-mail: markovng@tpu.ru

***Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.***

*Поступила в редакцию 06.11.2021; принята к публикации 30.08.2022*

*Received 06.11.2021; accepted for publication 30.08.2022*