

ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ АНАЛИЗА КОМПОНЕНТНОГО СОСТАВА ГАЗОВЫХ СМЕСЕЙ

Брокарев И.А.¹

(РГУ нефти и газа (НИУ) имени И.М. Губкина)

Приведена постановка задачи анализа компонентного состава природного газа. Предложено применение статистического метода для решения поставленной задачи анализа компонентного состава газа. Описаны основные этапы разработки статистической модели для анализа компонентного состава природного газа. Приведены результаты корреляционного анализа, проводимого для выбора входных и выходных параметров для статистической модели. Описаны основные статистические модели, используемые для решения задач анализа компонентного состава газовых смесей. Для решения исследуемой задачи предложено применить искусственные нейронные сети и алгоритм обучения Левенберга – Марквардта. Приведено описание алгоритма обучения Левенберга – Марквардта с учетом возможных модификаций данного алгоритма. Описана архитектура предлагаемой нейросетевой модели для решения задачи анализа компонентного состава газа. Приведены диапазоны газовых смесей, используемые в обучающей и тестовой выборках. Приведены точностные характеристики работы предлагаемой модели. На основе рассчитанных точностных характеристик модели сделан вывод об адекватности применения выбранной архитектуры нейросетевой модели. Приведены результаты использования предлагаемой нейросетевой модели для получения искомого компонентного состава по измерениям физических параметров газа. Приведены дальнейшие направления исследования в области разработки предлагаемого метода анализа компонентного состава природного газа.

Ключевые слова: нейросетевой анализ, алгоритм Левенберга – Марквардта, анализ компонентного состава, природный газ.

1. Введение

В настоящее время искусственный интеллект используется в решении множества прикладных задач. Одним из важных направлений искусственного интеллекта являются статистические модели. Среди большого разнообразия статистических моделей выделяют искусственные нейронные сети (ИНС) – математические модели, построенные по принципам организации и функционирования биологических нейронных сетей.

¹ Иван Андреевич Брокарев, аспирант (brokarev.i@gubkin.ru).

Одним из главных преимуществ нейронных сетей перед традиционными алгоритмами является возможность их обучения для совершенствования способности ИНС выявлять сложные зависимости между входными и выходными данными.

В связи с модернизацией и интеллектуализацией нефтегазовой промышленности, а также со сложностью решения промышленных задач традиционными методами в данной отрасли существует большое число примеров применения нейросетей, в частности, для прогнозирования работы скважин [8, 16], при обработке экспериментальных данных процесса коксования [5], прогнозирования эффективности технологических процессов [3, 6, 7], оптимизации процесса добычи нефти [11], а также в большом количестве других направлений [9].

Одной из перспективных задач, в которых возможно применение ИНС, является задача анализа компонентного состава природного газа. Смещение газов из разных месторождений и источников в процессе транспортировки и хранения приводит к изменениям компонентного состава природного газа на местах эксплуатации или учета природного газа. Поскольку компонентный состав газа определяет показатели качества газа, для правильной и оптимальной эксплуатации, учета и транспортировки природного газа трубопроводным транспортом становятся необходимыми измерение или расчет определенных свойств природного газа. Это делает актуальной задачу определения компонентного состава газа в реальном масштабе времени, для решения которой в настоящее время разрабатываются корреляционные методы анализа качества газа [4, 10]. В такого рода методах зачастую используются статистические модели, в частности ИНС, для определения искомых свойств и состава природного газа по измерениям физических параметров газа.

В данной статье рассмотрена предлагаемая нейросетевая модель для определения компонентного состава газа.

2. Нейросетевая модель для решения задачи анализа компонентного состава газовых смесей

Разработка модели для решения задачи анализа компонентного состава газовых смесей включает в себя ряд последова-

тельных этапов, а именно: выбор данных для обучения модели, выбор архитектуры модели, выбор метода обучения модели, оценка точности работы модели.

Первый этап заключался в выборе входных и выходных данных для модели. Критериями для выбора входных физических параметров являлись наличие взаимосвязи между данным параметром газа и компонентным составом, а также возможность измерения данного параметра коммерчески доступными и относительно недорогими измерительными приборами [13, 18]. Для определения взаимосвязи параметров и компонентного состава природного газа был проведен корреляционный анализ для проверки взаимозависимостей внутри множеств входных и выходных параметров и связи их между собой с целью устранения возможной мультиколлинеарности параметров – линейной взаимосвязи двух или нескольких переменных. Мультиколлинеарность может привести к нежелательным последствиям, так как при этом оценки параметров становятся ненадежными, т.е. возрастает стандартная ошибка, и становится невозможным определить изолированное влияние факторов на результирующий показатель. В качестве входных параметров необходимо использовать те показатели, которые имеют высокую корреляцию с выходными параметрами и низкую корреляцию между собой. По результатам корреляционного анализа, которые показаны в таблице 1, в качестве входных параметров были выбраны концентрация диоксида углерода, скорость звука и коэффициент теплопроводности. В качестве выходных параметров использовались подлежащие определению концентрации компонентов газовой смеси. На рис. 1 показаны результаты корреляционного анализа для параметров, которые не были выбраны в качестве входных по причине низкой корреляции (скорость звука на данном графике показана в качестве референсного параметра для сравнения). На рис. 2 показаны результаты корреляционного анализа для параметров, которые были выбраны в качестве входных по причине высокой корреляции.

Необходимо отметить, что перед обучением модели данные подвергаются кросс-валидации, т.е. перекрестной проверке – методу оценки аналитической модели и её поведения на независимых данных. Данная процедура заключается в следующем:

имеющиеся в наличии данные разбиваются на k частей, затем на $k - 1$ частях данных производится обучение модели, а оставшаяся часть данных используется для тестирования модели. Процедура повторяется k раз; в итоге каждая из k частей данных используется для тестирования. В результате получается оценка эффективности выбранной модели с наиболее равномерным использованием имеющихся данных. Для данной задачи выборка размером $N = 711480$ элементов разбивалась на $k = 10$ частей. Также обучающая выборка перед тренировкой подвергалась процедуре нормализации с целью улучшения результатов прогноза исследуемой нейросетевой моделью. Диапазоны входящих в обучающую выборку компонентов показаны в таблице 2.

Таблица 1. Коэффициенты корреляции входных физических параметров по отношению к содержанию компонентов газа

	Скорость звука, с, м/с	Коэффициент теплопроводности, χ , мВт/(м·К)	Содержание метана, %	Содержание азота, %	Содержание диоксида углерода, %	Содержание этана, %	Содержание пропана, %
Скорость звука, с, м/с	1	0,99	0,93	-0,24	-0,61	-0,33	-0,67
Коэффициент теплопроводности, χ , мВт/(м·К)	0,99	1	0,91	-0,15	-0,53	-0,45	-0,71
Содержание метана, %	0,93	0,91	1	-0,52	-0,52	-0,48	-0,49
Содержание азота, %	-0,24	-0,15	-0,52	1	0	0	0
Содержание диоксида углерода, %	-0,61	-0,53	-0,52	0	1	0	0
Содержание этана, %	-0,33	-0,45	-0,48	0	0	1	0
Содержание пропана, %	-0,68	-0,75	-0,49	0	0	0	1

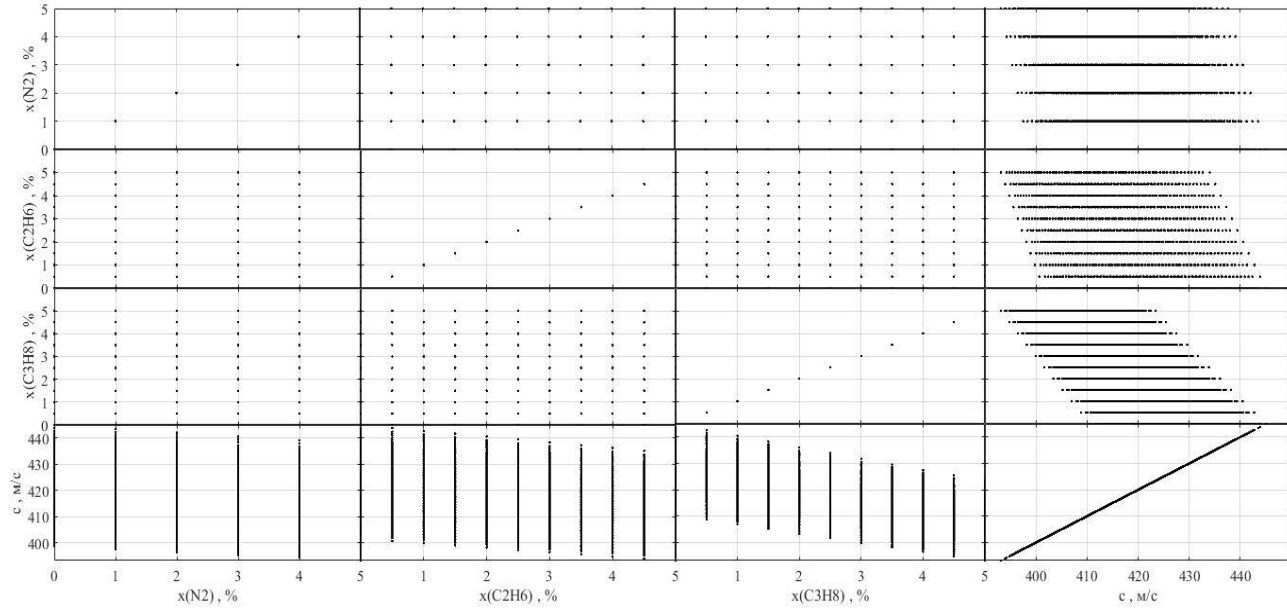


Рис. 1. Результаты корреляционного анализа для невыбранных параметров

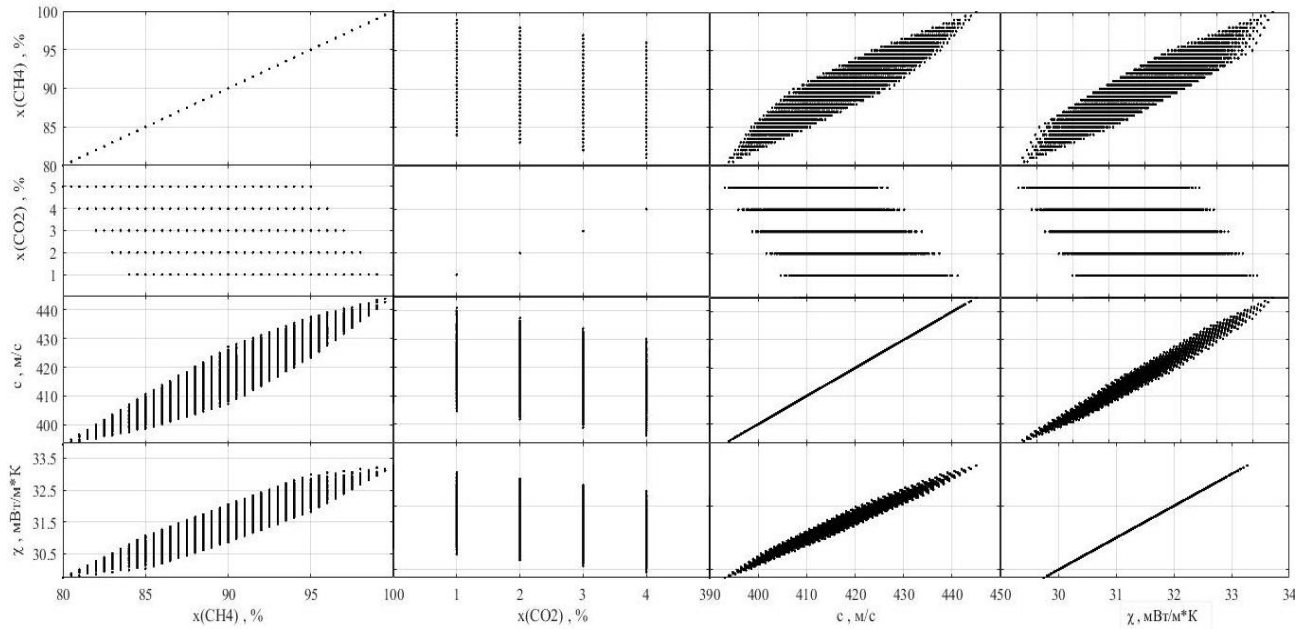


Рис. 2. Результаты корреляционного анализа для выбранных параметров

Таблица 2. Диапазоны газовых смесей для обучающей выборки

Компонент	Молярная доля, %
Метан	80 – 100
Этан	0 – 5
Пропан	0 – 5
Диоксид углерода	0 – 5
Азот	0 – 5

В качестве статистических моделей для решения задачи анализа качества газа могут быть использованы различные модели:

- Многопараметрическая линейная регрессия может служить референсной моделью. Она используется для получения результата, который будет взят для сравнения точности других моделей с регрессионной. Для данной модели необходимо отдельно производить расчет для каждого компонента исследуемой газовой смеси, что неэффективно при наличии других моделей.
- Модель на основе метода опорных векторов (support vector machine – SVM). Это набор схожих алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа. Основная идея метода – перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Данная модель не была выбрана по причине наличия существенных недостатков, в частности сложной интерпретации параметров модели и применения только для решения задач с двумя классами.
- Нейросетевая модель.

Нейросетевую модель предлагается использовать в качестве основной статистической модели в рассматриваемой задаче. Для решения поставленной задачи в качестве архитектуры модели использовалась трехслойная сеть (многослойный персептрон), включающая в себя два нейрона во входном слое – по числу компонент входного вектора, одиннадцать нейронов во втором слое и четыре нейрона в выходном слое – по числу компонентов

выходного вектора. Входной слой представляет собой вектор $X = \{x_0, x_1, x_2\}$. Нейроны каждого следующего слоя связаны с нейронами предыдущего слоя, причем каждый входной сигнал имеет определенный вес, который в данном случае одинаков для всех входных нейронов в силу равной важности всех входных величин. У каждого нейрона есть функция активации, аргументом которой является входной сигнал нейрона. У нейронов скрытого слоя функция активации сигмоидальная в виде гиперболического тангенса, у нейронов выходного слоя – линейная. В качестве обучающего алгоритма выбран алгоритм Левенберга – Марквардта (Levenberg – Marquardt) [2]. Данный алгоритм используется для оптимизации параметров нелинейных регрессионных моделей. Отметим, что в данном алгоритме критерием оптимизации является среднеквадратичная ошибка модели на обучающей выборке. Основная идея алгоритма состоит в том, чтобы достичь искомого локального оптимума путем приближения к нему заданных начальных значений параметров.

Пусть заданы регрессионная выборка, т.е. множество пар свободной переменной X и зависимой переменной Y , и регрессионная модель – непрерывно дифференцируемая функция F . Требуется найти такое значение вектора параметров W , в котором функция ошибки F_ε достигает своего локального минимума:

$$(1) \quad F_\varepsilon = \sum_{i=1}^N (Y_i - F(X_i, W))^2 .$$

На первой итерации алгоритма задается начальный вектор параметров W_0 . На каждой следующей итерации вектор заменяется на вектор $W_0 + \Delta W$. Для оценки приращения ΔW используется приближение функции F :

$$(2) \quad F(W_0 + \Delta W, X) - F(W_0, X) = J * \Delta W ,$$

где J – якобиан функции F .

Приращение ΔW в точке, доставляющий минимум функции F_ε , равно нулю, поэтому для нахождения последующего значения приращения ΔW необходимо приравнять нулю вектор частных производных функции F_ε по W . Для этого F_ε необходимо представить в следующем виде и раскрыть скобки:

$$(3) \quad F_\varepsilon = |Y - F(W + \Delta W)|^2.$$

$$(4) \quad F_\varepsilon = (Y - F(W + \Delta W))^T * (Y - F(W + \Delta W)) + Y^T * Y - 2 * Y^T * F(W + \Delta W).$$

Дифференцируя данное выражение по W , получим выражение

$$(5) \quad \frac{\partial F_\varepsilon}{\partial W} = J^T * J * \Delta W - J^T * (Y - F(W)).$$

Приравняв частную производную к нулю, можем получить значение ΔW :

$$(6) \quad \Delta W = (J^T * J)^{-1} * J^T * (Y - F(W)).$$

Для рассматриваемого алгоритма важным является следующий параметр – число обусловленности матрицы, которое показывает, насколько матрица близка к матрице неполного ранга (для квадратных матриц – к вырожденности). Учитывая, что число обусловленности матрицы $J^T * J$ равно квадрату числа обусловленности матрицы J , то матрица $J^T * J$ может оказаться вырожденной. По этой причине в данном алгоритме введен параметр регуляризации $\lambda \geq 0$. Данный параметр выбирается на каждой итерации алгоритма. Учитывая параметр регуляризации, выражение для ΔW примет вид:

$$(7) \quad \Delta W = (J^T * J + \lambda * E)^{-1} * J^T * (Y - F(W)),$$

где E – единичная матрица.

Возможна модификация данного метода [14]: параметр регуляризации умножается на матрицу D – диагональную матрицу с элементами, совпадающими с диагональными элементами матрицы $J^T * J$. Данный подход применяется для снижения влияния параметра регуляризации на значение ΔW . Необходимо, однако, отметить, что недостатком рассматриваемого алгоритма является низкая сходимость при постоянной длине шага. Эта проблема решается путем ввода коэффициента K , который определяет длину шага и увеличивает сходимость метода. Учитывая две описанные поправки, выражение для ΔW примет следующий вид:

$$(8) \quad \Delta W = K * (J^T * J + \lambda * D)^{-1} * J^T * (Y - F(W)).$$

Значение вектора W на последней итерации алгоритма считается искомым. Оно достигается либо в случае если рассчитан-

ное приращение ΔW меньше заданного значения ΔW_{set} , либо если значение функции ошибки F_e меньше заданной величины $F_{e_{set}}$ при данном значении вектора W .

Архитектура используемой нейросетевой модели показана на рис. 3. Количество нейронов во входном слое $n = 3$ для случая, когда в векторе входных параметров присутствуют концентрация диоксида углерода, скорость звука и теплопроводность. Количество нейронов в скрытом слое $k = 11$ для данной конкретной модели и выбрано в соответствии с анализом различных моделей. Количество нейронов в выходном слое $m = 4$ для случая пятикомпонентной газовой смеси. W_i и b_i – весовые коэффициенты и коэффициенты смещения для скрытого слоя, W_j и b_j – для выходного слоя.

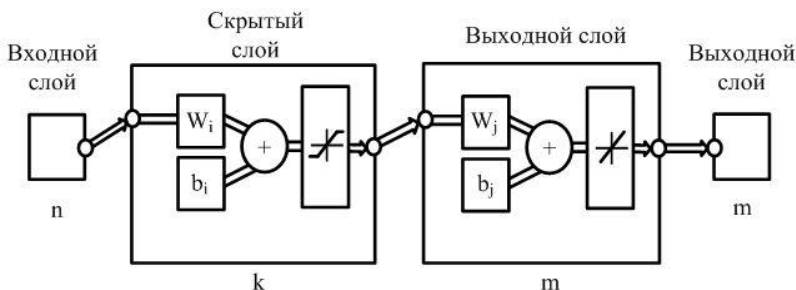


Рис. 3. Архитектура предлагаемой нейросетевой модели

Следующий шаг заключается в обучении созданной сети на сформированных заранее теоретических данных с использованием рассмотренного выше алгоритма Левенберга – Марквардта. Перед этим необходимо задать критерий окончания обучения, например, максимальное значение отклонения, при котором обучение будет считаться законченным. В нашем случае для получения оценки вычисляется сумма квадратичных отклонений выходов сети от истинных значений. Также при обучении модели одним из критериев окончания обучения может выбираться максимальное количество циклов обучения, но, как видно из рис. 4, выбор такого критерия приводит к неэффективному увеличению времени обучения.

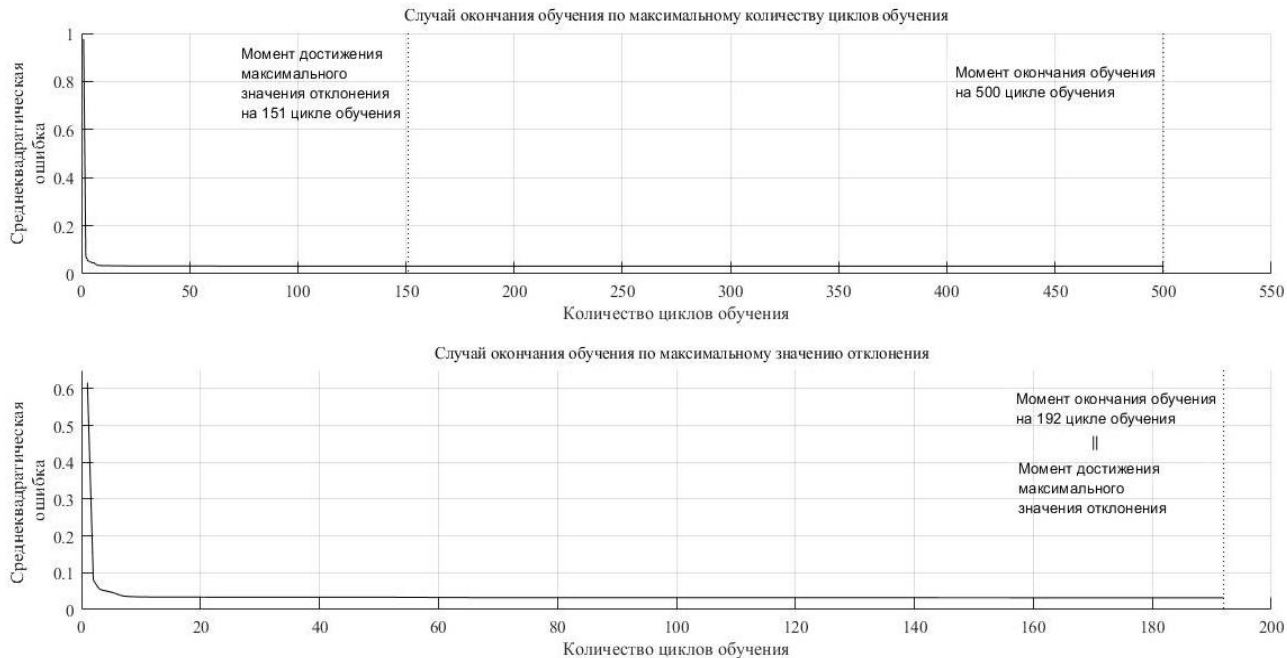


Рис. 4. Случаи окончания обучения для разрабатываемой нейросетевой модели

После успешного окончания обучения полученная нейросетевая модель тестируется на данных, которые не участвовали в тренировке модели. Затем для оценки точности работы модели рассчитываются следующие точностные параметры:

- Максимальная абсолютная погрешность:

$$(9) \Delta Y = \max[|Y_{out} - Y_{target}|],$$

где Y_{out} – значения выхода нейросети, Y_{target} – целевые значения параметра.

- Средняя абсолютная погрешность:

$$(10) \Delta Y = \text{avg}[|Y_{out} - Y_{target}|].$$

- Максимальная относительная погрешность (в %):

$$(11) \delta_Y = \max\left[\frac{|Y_{out} - Y_{target}|}{Y_{target}}\right].$$

- Средняя относительная погрешность (в %):

$$(12) \delta_Y = \text{avg}\left[\frac{|Y_{out} - Y_{target}|}{Y_{target}}\right].$$

- Среднеквадратическое отклонение:

$$(13) MSE = \frac{\sum_{i=1}^n (Y_{out} - Y_{target})^2}{n}.$$

• Коэффициент детерминации – параметр, показывающий насколько исследуемая модель соответствует данным. Равенство коэффициента детерминации единице означает, что данные в точности описываются рассматриваемой моделью:

$$(14) R^2 = 1 - \frac{\sum_{i=1}^n (Y_{out} - Y_{target})^2}{\left(\sum_{i=1}^n Y_{out} - \frac{\sum_{i=1}^n Y_{out}}{n}\right)^2}.$$

По перечисленным параметрам делается вывод о возможности использования модели для решения исследуемой задачи. В случае если модель обладает низкими точностными характе-

ристиками, повторяется этап обучения модели либо применяется другая архитектура модели.

3. Применение разработанной нейросетевой модели

По результатам проведенного исследования была получена нейросетевая модель для расчета компонентного состава газа по измеряемым физическим параметрам. Для проверки точности работы предложенного метода было проведено исследование на тестовой выборке данных. Данная выборка включала данные, которые не участвовали в обучении нейросетевой модели. Тестовая выборка включала 4356 газовых смесей, диапазоны входящих в нее компонентов показаны в таблице 3. Расчет физико-химических параметров для смоделированной выборки производился с помощью программного комплекса NIST REFPROP [17]. При расчете скорости звука использовалось уравнение состояния газа GERG-2008 для стандартных условий по температуре и давлению. Теплопроводность газовой смеси рассчитывалась на основе моделей для индивидуальных компонентов и расширенной модели соответствующих состояний, реализованных в NIST REFPROP.

Таблица 3. Диапазоны газовых смесей для тестовой выборки

Компонент	Молярная доля, %
Метан	80,5 – 99,5
Этан	0,5 – 4,5
Пропан	0,5 – 4,5
Диоксид углерода	0,5 – 4,5
Азот	0,5 – 4,5

Точностные характеристики результата предсказания моделью компонентного состава природного газа на тестовой выборке показаны в таблице 4 и на рис. 5. Отклонение для диоксида углерода принималась равным нулю, так как содержание в газе данного компонента в рассматриваемой задаче считается известным из-за наличия доступной технологии для его измерения, а именно технологии недисперсионного поглощения ин-

фракрасного излучения (NDIR) [12]. Расчет параметров, а также построение, обучение и тестирование нейросетевой модели проводились в пакете Matlab 2018a с использованием функционала Deep Learning Toolbox [15].

Таблица 4. Точность предсказания компонентного состава газа разработанной нейросетевой моделью

Компонент	Максимальное абсолютное отклонение, %	Среднее абсолютное отклонение, %	Среднеквадратическое отклонение	Коэффициент детерминации
Метан	0,75	0,32	0,34	0,9999
Этан	0,37	0,15	0,17	0,9999
Пропан	0,25	0,11	0,13	0,9999
Азот	0,21	0,09	0,10	0,9999
Диоксид углерода	0	0	0	1

4. Выводы и дальнейшие исследования

Предложенная нейросетевая модель использовалась для прогноза состава газа по значениям его физико-химических параметров. Параметры природного газа, выбранные в качестве входных для модели, можно измерять с помощью коммерчески доступного измерительного оборудования. Анализ адекватности применения разработанной модели показал, что методологическая ошибка при определении содержания метана, этана, пропана и азота удовлетворяет требованиям газовой промышленности [1]. Дальнейшие исследования в данной области заключаются в проверке метода в лабораторных и промышленных условиях. Также планируется реализовать модель с использованием более сложной архитектуры для улучшения точности ее работы.

Автор выражает благодарность международным проектам сотрудничества между университетами BRISK II TA и Erasmus+ 2017-1-SE01-KA107-034292 Staff Mobility за предоставленную возможность для проведения данного исследования.

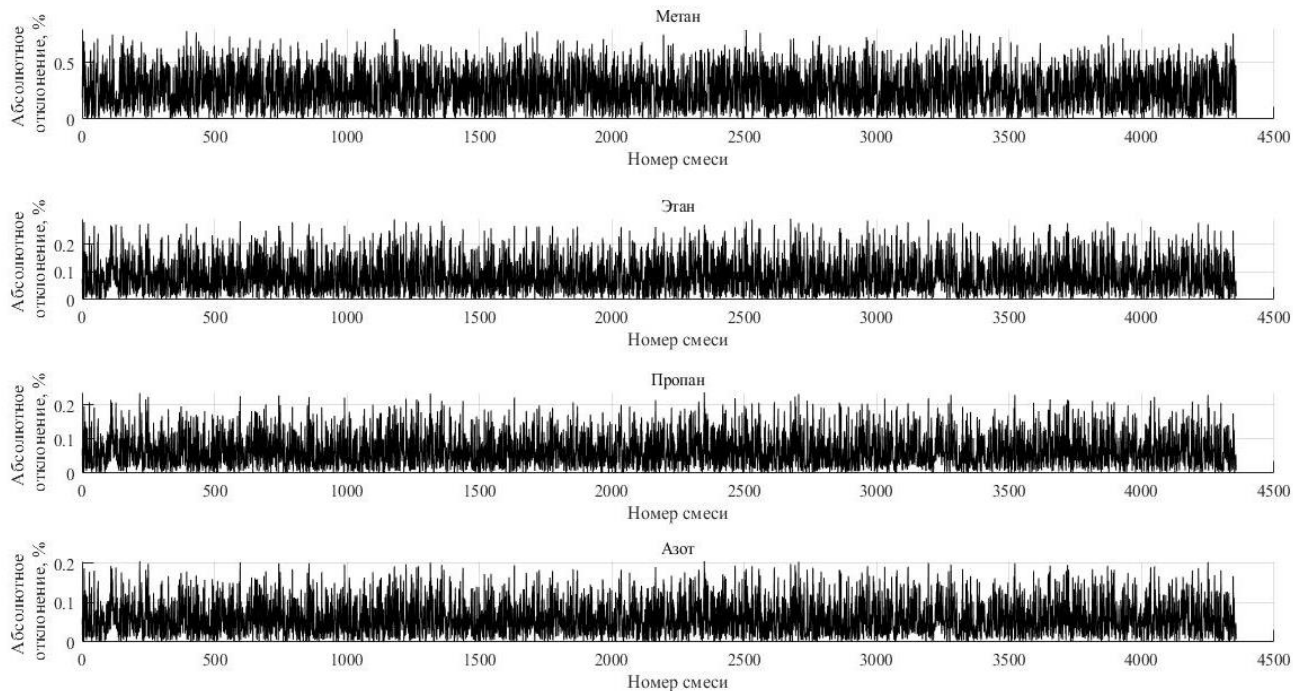


Рис. 5. Абсолютное отклонение результата предсказания моделью компонентного состава природного газа на тестовой выборке

Литература

1. *ГОСТ 5542-2014. Газ горючий природный промышленного и коммунально-бытового назначения.* – М.: Стандартинформ, 2015. – С. 7–8.
2. ДЕМИДЕНКО Е.З. *Оптимизация и регрессия.* – М.: Наука, 1989. – С. 223–228.
3. КЕЛЛЕР Ю.А. *Разработка искусственных нейронных сетей для предсказания технологической эффективности от выравнивания профиля приемистости // Известия Томского политехнического университета. Информационные технологии.* – 2014. – №5. – С. 60–65.
4. КОТУРБАШ Т.Т., БРОКАРЕВ И.А. *Метод определения свойств и состава природного газа по измерениям его физических параметров // Датчики и системы.* – 2018. – №6. – С. 43–50.
5. ЛЕОНОВ М.Г., ЖИРНОВ Б.С. *Обработка экспериментальных данных процесса коксования с помощью нейронных сетей // Нефтегазовое дело.* – 2014. – №2. – С. 151–165.
6. МАНДРИК И.Э., ШАХВЕРДИЕВ А.Х., СУЛЕЙМАНОВ И.В. *Оценка и прогноз нефтеотдачи на основе моделирования нейронными сетями // Нефтяное хозяйство.* – 2005. – №10. – С. 31–34.
7. ПЧЕЛЬНИКОВ И.В., БОРХОВИЧ С.Ю., НАТАРОВ А.Л. *Перспективы прогнозирования эффективности ГТМ на основе нейросетевого моделирования // Нефть. Газ. Новации.* – 2016. – №4. – С. 37–40.
8. СОЛОМАТИН Г.И., ЗАХАРЯН А.З., АШКАРИН Н.И. *Прогнозирование работы скважин с помощью искусственных нейронных сетей // Нефтяное хозяйство.* – 2002. – №10. – С. 92–96.
9. ASHENA R., THONHAUSER G *Application of Artificial Neural Networks in Geoscience and Petroleum Industry // Artificial Intelligent Approaches in Petroleum Geosciences.* – 2015. – P. 127–166.

10. DÖRR H., KOTURBASH T., KUTCHEROV V. *Review of impacts of gas qualities with regard to quality determination and energy metering of natural gas* // Measurement Science and Technology. – 2019. – Vol. 30, No. 2. – P. 1–20.
11. JIMENEZ DE LA CRUZ G., RUZ-HERNANDEZ J., SHELOMOV E., SALAZAR-MENDOZA R. *Optimization of an Oil Production System using Neural Networks and Genetic Algorithms* // Proc. of the Joint 2009 Int. Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference. – 2009. – P. 1815–1820.
12. KAUR J., ADAMCHUK V., WHALEN J., ISMAIL A. *Development of an NDIR CO₂ Sensor-Based System for Assessing Soil Toxicity Using Substrate-Induced Respiration* // Sensors. – 2015. – №15. – P. 4734–4748.
13. KOTURBASH T., BICZ A., BICZ W. *New instrument for measuring velocity of sound and quantitative characterization of binary gas mixtures composition* // Measurement Automation Monitoring. – 2016. – P. 254–258.
14. MARQUARDT D.W. *An algorithm for least squares estimation of non-linear parameters* // J. Soc. of Appl. Math. – 1963. – Vol. 2. – P. 431–441.
15. *Matlab Software*. – URL: <https://www.mathworks.com>.
16. MIRZAEI-PAIAMAN A., SALAVATI S. *The Application of Artificial Neural Networks for the Prediction of Oil Production Flow Rate* // Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. – 2012. – No. 34:19. – P. 1834–1843.
17. *REFPROP Software*. – URL: <https://www.nist.gov/srd/refprop>.
18. *Thermal Conductivity Gauge* [Electronic resource]. – URL: <http://www.xensor.nl>.

USING ARTIFICIAL NEURAL NETWORKS FOR SOLVING THE PROBLEM OF ANALYSIS OF THE COMPOSITION OF GAS MIXTURES

Ivan Brokarev, National University of Oil and Gas (Gubkin University), Moscow, Ph.D. Student (brokarev.i@gubkin.ru).

Abstract: The problem formulation of natural gas composition analysis is described. The statistical method is proposed to use for solving the problem of gas composition analysis. The main stages of the statistical model development for the natural gas composition analysis are described. The results of the correlation analysis for the selection of input and output parameters for the statistical model are presented. The main statistical models that used for solving of the problem of gas mixtures composition analysis are shown. The artificial neural networks and Levenberg – Marquardt algorithm are used within the study. The description of the Levenberg – Marquardt training algorithm is given taking into account possible modifications of the algorithm. The architecture of the proposed neural network is described. The gas mixtures ranges that used in the training and test samples are given. The accuracy characteristics of the proposed model are given. It was concluded that the chosen neural network model architecture is adequate, based on the calculated accuracy characteristics of the model. The results of predicting the gas mixtures composition by measurements of gas physical parameters are shown. Further research directions in the development of the proposed method for analyzing the natural gas composition are given.

Keywords: artificial neural network analysis, Levenberg – Marquardt algorithm, analysis of the composition, natural gas.

УДК 519.6
ББК 30.1

*Статья представлена к публикации
членом редакционной коллегии В.М. Вишневым.*

*Поступила в редакцию 07.03.2019.
Опубликована 31.07.2019.*