

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН



КазҰТЗУ ХАБАРШЫСЫ

ВЕСТНИК КАЗНИТУ

VESTNIK KazNRTU

№ 5 (141)

<i>Болатов Н.К., Оразов О.Ш., Болатова Ж.Т., Ахметжанова А.Б.</i>	
БИОГАЗ ҚОНДЫРҒЫСЫНЫң ТӘЖІРИБЕЛІК ҮЛГІСІНДЕ АЭРОБТЫ ЖӘНЕ АНАЭРОБТЫ АФЫНДАРДЫҢ ТИМДІ АРАҚАТЫНАСЫН АНЫҚТАУ	248
<i>Болатов Н.К., Қасымжанова Қ.С., Болатова Ж.Т., Мукашев А.А.</i>	
БИОМАССАНЫ АРАЛАСТЫРУДЫҢ ЖӘНЕ АЙНАЛЫМЫНЫң ОҢТАЙЛЫ ЖҮЙЕСІ БАР ЭКСПЕРИМЕНТТІК БИОГАЗ ҚОНДЫРҒЫСЫНЫң ШЫҒЫС ПАРАМЕТРЛЕРІН АНЫҚТАУ.....	255
<i>Болатов Н.К., Қасымжанова Қ.С., Оразов О.Ш., Болатова Ж.Т.</i>	
БИОМАССАНЫ АРАЛАСТЫРУДЫҢ ЖӘНЕ АЙНАЛЫМЫНЫң ОҢТАЙЛЫ ЖҮЙЕСІ БАР ЭКСПЕРИМЕНТТІК БИОГАЗ ҚОНДЫРҒЫСЫНДА ПРОЦЕСТИҢ ТҮРАҚТЫЛЫҒЫН АНЫҚТАУ...	263
<i>Болатов Н.К., Аятхан М.А., Болатова Ж.Т., Ахметжанова А.Б.</i>	
БИОГАЗ ҚОНДЫРҒЫСЫНЫң ТӘЖІРИБЕЛІК ҮЛГІСІН ЭКСПЕРИМЕНТТІК ЖӘНЕ ИНЖЕНЕРЛІК СЫНАУ.....	271
<i>Дайырбаева Э.Н., Мурzin Ф.А., Тойгожинова А.Ж., Липская М.А.</i>	
СҮРЕТТЕРДІ ӨҢДЕУДЕ СТРИП-ӘДІСТІ ПАЙДАНАЛУ ЖОЛДАРЫ МЕН НӘТИЖЕЛЕРІ.....	279
<i>Сәндібек Д.</i>	
ИНТЕЛЛЕКТУАЛДЫҚ ДИЗАЙНДЫҢ ҮЛГІЛЕРІ МЕН ӘДІСІН ДАМЫТУ АЭРОПОРТТЫҚ АГЕНТТІК ЖҮЙЕСІ.....	284
<i>Ескокжанова Н.Г., Нұржасаубаев М.М., Тилемисова А.А., Нуралина А.К.</i>	
ЛОГИСТИКАНЫң ДАМУЫНЫң ҚАЗІРГІ КӨЛІКТІК ТЕНДЕНЦИЯЛАРЫ.....	289
<i>Ескокжанова Н.Г., Нұржасаубаев М.М., Тилемисова А.А., Саукенова И.К.</i>	
ҚАЗІРГІ ШАРТТАРДА КӨЛІК МАРШРУТТАРЫНЫң ЛОГИСТИКАСЫ.....	292
<i>Жабаев Т.Р., Тукеев У.А.</i>	
МӘТИНДІ РЕТТІЛІК МОДЕЛЬДЕРІНЕ ТІЗБЕКТЕЙ ҚОЛДАNU АРҚЫЛЫ ЖЕҢІЛДЕТУ.....	294
<i>Жақсылық Р.Ұ.</i>	
МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАNU АРҚЫЛЫ СӨЙЛЕУ ЭМОЦИЯЛАРЫН ТАНУ ӘДІСТЕРІ.....	298
<i>Досжанова А.А., Зтелеу Ш.С., Заманғалиева Ш.Ш.</i>	
КӨЛІКТІК ИНФРАҚҰРЫЛЫМЫНДА КИБЕРШАБУЫЛДЫ ТАНУДЫҢ АҚПАРАТТЫҚ ҚАУІПСІЗДІК МОДЕЛІ МЕН АЛГОРИТІМІ.....	302
<i>Айдарова С., Шарипова А., Исаева А., Султанбеккызы Д., Исахов М., Бабаев А.</i>	
НАТРИЙ КАЗЕИНАТЫ МЕН ХИТОЗАННЫң МАЙМЕН ШЕКАРАСЫНДА ФАЗАРАЛЫҚ КЕРІЛУІНІң ДИНАМИКАЛЫҚ ЕРЕКШЕЛІКТЕРІН ЗЕРТТЕУ	306
<i>Исембергенов Н.Т., Абдиқул Т.</i>	
КҮН ЭЛЕКТРОСТАНЦИЯСЫНА АРНАЛҒАН ТРАНЗИСТОРЛЫҚ ИНВЕРТОРДЫ ЗЕРТТЕУ.....	312
<i>Исембергенов Н.Т., Искакова А.Ж.</i>	
ИНВЕРТОРЛЫҚ КЕРНЕУ СТАБИЛИЗАТОРЫН ЗЕРТТЕУ	316
<i>Исраилова С., Муханова А., Исмаилова А., Маханов М.</i>	
МИТАЦИЯЛЫҚ МОДЕЛЬДЕУДІ ҚОЛДАNUЫМЕН ТЕРМОБОКСТЫң САЛҚЫНДАТУ ЭЛЕМЕНТИНІң ТЕМПЕРАТУРАСЫН ТҮРАҚТАНДЫРУ БИЗНЕС-ПРОЦЕСІН ЗЕРТТЕУ	321
<i>Кочеров Е.Н., Мақұлбекова Г.О., Кенжалиева Г.Д., Мамитова А.Д., Бегимбетова А.С.</i>	
ҚЫҢҒЫРАҚ-КЕЛЕС БЕНТОНИТТІ САЗДАРЫНЫң КЕУЕКТЕНУІНЕ ҚОҢЫР КӨМІРДІң ЫҚПАЛАЫ.....	325
<i>Казанғапова Б.А.</i>	
КӘСПОРЫНДАРДЫҢ АҚПАРАТТЫҚ ҚАУІПСІЗДІГІН ҚАМТАМАСЫЗ ЕТУ БОЙЫНША ӘРТҮРЛІ АҚПАРАТТЫҚ ЖҮЙЕЛЕРДІ САЛЫСТЫРМАЛЫ ТАЛДАУ	330
<i>Калижанова А.У., Кемельбекова И.М., Нұрсейт Н.Н., Мустафаева А.А.</i>	
БОЛЖАУ МОДЕЛЬДЕРІ МЕН ӘДІСТЕРІН ЗЕРТТЕУ	333
<i>Қамен Б.Қ.</i>	
ҚАЗАҚСТАНДАҒЫ БІЛІМ БЕРУ САЛАСЫНДАҒЫ РОБОТОТЕХНИКА: ТӘЖІРИБЕ ЖӘНЕ ДАМУ БОЛАШАҒЫ.....	341
<i>Кашкимбаева Н.М., Калдарова М.Ж.</i>	
АУЫЛШАРУАШЫЛЫҚ СЕНСОРЛАРЫН БАҚЫЛАУ ЖҮЙЕСІН ДАМЫТУ.....	343
<i>Жанибек К.Г.</i>	
ӨНДІРІСТЕГІ АҚПАРАТТЫҚ ҚАУІПСІЗДІКТІ ҚАМТАМАСЫЗ ЕТУ САЯСАТЫ.....	347
<i>Кемельбеков И.М., Смагулов Б.М., Сергазин А.Ә., Кошабаев Д.А.</i>	
PUSH-НОТИФИКАЦИЯЛАР АРҚЫЛЫ АУАДАҒЫ ЗИЯНДЫ ЗАТТАРДЫ АНЫҚТАУҒА ЖӘНЕ ТАРАТУҒА АРНАЛҒАН ЗИЯТКЕРЛІК ЖҮЙЕНІ ҚҰРУ.....	352
<i>Кенжебек Е.Ғ., Иманкулов Т.С., Ахмед-Заки Д.Ж.</i>	
ПОЛИНОМДЫ РЕГРЕССИЯ ӘДІСІМЕН МҰНАЙ ӨНДІРУДІ БОЛЖАУ	358

УДК 004.85

Y.G. Kenzhebek¹, T.S. Imankulov¹, D.Zh. Akhmed-Zaki²

(¹al-Farabi Kazakh National University, Almaty, Kazakhstan,

²University of International Business, Almaty, Kazakhstan,

E-mail: kenzhebekyerzhan@gmail.com)

PREDICTIONS OF OIL PRODUCTION BY THE POLYNOMIAL REGRESSION METHOD

Annotation. This article presents the work of predicting oil production using machine learning methods. As a machine learning method, a multivariate linear regression algorithm with polynomial properties was implemented. The synthetic dataset was obtained using the Buckley-Leverett mathematical model, which is used to calculate hydrodynamics and determine the saturation distribution in oil production problems. Various combinations of parameters of the oil production problem were selected, where porosity, oil phase viscosity and absolute rock permeability were taken as input parameters for machine learning. And as an output parameter, the value of the oil recovery factor was chosen. To test the polynomial regression algorithm, 10,086 data pairs were used. The Python programming language was used as a runtime for machine learning. Various degrees of polynomial regression were deployed and tested, and it was also found that for our synthetic data, the quadratic polynomial model is quite well trained and perfectly predicts the value of the oil recovery coefficient. To solve the overfitting problem, a regularization of the form L1, known as the Lasso regression method was applied. For a quadratic polynomial regression model, the coefficient of determination R^2 is 0.96, which is a fairly good result for the test data.

Key words: enhanced oil recovery, machine learning, regression method, polynomial regression, regularization.

Е.Ф. Кенжебек¹, Т.С. Иманкулов¹, Д.Ж. Ахмед-Заки²

(¹Казахский национальный университет имени аль-Фараби, Казахстан,

²Университет международного бизнеса, Казахстан,

E-mail: kenzhebekyerzhan@gmail.com)

ПРЕДСКАЗАНИЕ ДОБЫЧИ НЕФТИ МЕТОДОМ ПОЛИНОМИАЛЬНОЙ РЕГРЕССИИ

Аннотация. В данной статье представлена работа предсказание добычи нефти с использованием методов машинного обучения. В качестве метода машинного обучения был реализован алгоритм многомерной линейной регрессии с полиномиальными свойствами. Синтетический набор данных был получен с помощью математического моделья Баклея-Леверетта, которая используется для расчета гидродинамики и определения распределения насыщенности в задачах нефтедобычи. Выбраны различные комбинации параметров задачи добычи нефти, где в качестве входных параметров для машинного обучения были взяты пористость, вязкость нефтяной фазы и абсолютная проницаемость породы. А в качестве выходного параметра был выбран значение коэффициента нефтеотдачи. Для тестирования алгоритма полиномиальной регрессии была использована 10086 пар данных. В качестве среды исполнения для машинного обучения был применен язык программирования Python. Развернута и протестированы разные степени полиномиальной регрессии, а также было выявлено, что для наших синтетических данных квадратичная полиномиальная модель довольно хорошо обучается и отлично прогнозирует значение коэффициента нефтеотдачи. Для решения проблемы переобучения была применена регуляризация вида L1, известная как метод регрессии Лассо. Для квадратичной модели полиномиальной регрессии коэффициент детерминации R^2 составляет 0.96, что является довольно хорошим результатом для тестовых данных.

Ключевые слова: повышение нефтеотдачи, машинное обучение, метод регрессии, полиномиальная регрессия, регуляризация.

Введение

Есть много работ связанных с улучшением добычи нефти с использованием методов машинного обучения. В одном из таких работ [1] авторы выясняли, что применение алгоритмов машинного обучения могут оказаться более производительными по сравнению с традиционными вычислениями на регулярной сетке. А так же в данной работе описывается подход к созданию прокси-модели [2, 3] на основе методов машинного обучения, в частности была использована метод случайного леса [4].

В работе [5] рассматривается алгоритмы машинного обучения для оценки коэффициента добычи нефти с использованием комбинации инженерных и стратиграфических параметров. Для набора данных, состоявшийся из 30 параметров были применены модели линейной регрессии и метод опорных векторов. В результате, полученные данные были очень близкими к результатам

перекрестной проверки. Таким образом, авторы данной работы предполагают, что рассмотренные ими методы могут использоваться для прогнозирования добычи в дальнейшем.

Авторы работы [6, 7] рассмотрели применение искусственной нейронной сети (ANN) для прогнозирования дебита добычи нефти. Разработанная модель (ANN) в данной работе [6] прогнозирует дебит нефти с помощью трех параметров. Кроме того, точность модели сравнивались с некоторыми популярными корреляциями, следовательно, авторы утверждают, что разработанная модель отлично согласуется с фактическими данными измерений. А в работе [7] в качестве входного слоя нейронной сети были применены 143 набора данных из 6 параметров. Таким образом, с помощью искусственной нейронной сети типа FFNN авторы получили обнадеживающие результаты для одного из рассмотренного нефтяного месторождения в их исследовании.

В исследовании [8] рассмотрены различные методы машинного обучения для прогнозирования давления в забое скважины, дебита нефти и прогноза обводненности в многофазных производственных задачах. Набор данных в данном исследовании были получены с помощью симулятора пласта ECLIPSE. В этом исследовании авторы применили десять различных методов машинного обучения, кроме этого были учтены влияние многофазного потока и шума данных. Гребневая регрессия и метод регрессии опорных векторов показали лучшие результаты при любых уровнях шума в их исследовании.

Есть несколько работ связанных с применением методов машинного обучения для обработки данных из постоянных скважинных датчиков (PDG) [9]. Данные PDG очень часто являются шумными из-за эксплуатационных изменений, происходящие в скважине. Считается что современные PDG могут записывать данные каждую секунду, таким образом, после нескольких месяцев работы создаются очень большие данные, которые являются трудными к обработке. Следовательно, авторы хотели разработать надежный метод для обработки данных из постоянных скважинных датчиков. В работах Liu и Horne [10, 11] был применен простое ядро и подходы к анализу данных на основе метода сверточного ядра (СК) для интерпретации данных из постоянных скважинных датчиков. Авторы показали, что метод СК отлично удаляет шум, однако выяснилось, что данный алгоритм работает очень медленно [12, 13]. Авторы работы [14] также рассмотрели применение методов машинного обучения для интерпретации данных давления, дебита и температуры из постоянных скважинных датчиков. В данной работе были применены три метода машинного обучения, такие как линейная регрессия (LR), метод ядра и гребневая регрессия ядра. Кроме того, авторы работ показали, что машинное обучение может моделировать сгенерированные данные из PDG, даже когда физическая модель является сложной.

В работе [15] рассматривается использование методов машинного обучения для прогнозирования, существующих и новых скважин. Авторы построили искусственную нейронную сеть, который прогнозирует производительность добычи скважин, используя их собственную историю. Однако авторы не утверждают что прогнозирование с помощью искусственной нейронной сети (ANN) является заменой эмпирического или численного моделирования для прогнозирования добычи скважин. Работа предполагает, что прогнозирование ANN следует использовать для уверенности в методах прогнозирования основанных на данных. Существует другая работа, в котором были построены методы машинного обучения для предсказания добычи скважин Montney и Duvernay [16]. Были рассмотрены несколько методов ML, из которых самой точной моделью для их задачи была выявлена метод случайного леса. Данный метод дал авторам более высокую точность прогноза из-за отсутствия проблем over-fitting, а так же коэффициент детерминации R^2 составила 0.75 для Montney и 0.68 для Duvernay скважины.

Цель данной работы заключается в применении методов линейной и полиномиальной регрессии для предсказания добычи нефти. Сравнение результатов и точности предсказания для обоих методов.

Метод и теория

В данном разделе описывается процесс работы и используемые методы машинного обучения. А так же рассматривается генерация набора данных, обучение и тестирование.

Целью данного исследования является применение методов машинного обучения для прогнозирования добычи нефти. Синтетический набор данных были получены с помощью математической модели Баклея-Леверетта, которая используется для расчета гидродинамики и определения распределения насыщенности в задачах нефтедобычи.



Рис. 1. Процесс построения модели машинного обучения

Рисунок 1 описывает процесс построения модели машинного обучения в данном исследовании. В данной работе полученные синтетические данные из математической модели были разделены на обучающую и тестовую выборку. В исследовании в качестве входных параметров модели машинного обучения были взяты 4 параметра, и в качестве выходного параметра был взят коэффициент нефтеотдачи пласта. Набор данных подробно описывается в следующей главе.

Генерация набора данных

С помощью математической модели Баклея-Леверетта был получен синтетический набор данных. С помощью метода основанного на ансамбле сценариев были получены данные, где в качестве входных параметров были взяты различные комбинации параметров задачи добычи нефти: пористости, вязкости нефтяной фазы и абсолютной проницаемости породы (Таблица 1). А в качестве выходного параметра был выбран значение коэффициента нефтеотдачи пласта.

Таблица 1. Входные параметры

Признаки	Количество вариации
Пористость	41
Вязкость нефтяной фазы	41
Проницаемость	6

Таким образом, в данной работе количество пар выборки составляет $41 \times 41 \times 6 = 10086$. С помощью модели Баклея-Леверетта сформированы 6 пакетов синтетических данных для различных показателей проницаемости. Каждый пакет данных содержит значений вязкости, пористости и коэффициента нефтеотдачи.

В каждом запуске рассчитывается значение коэффициента нефтеотдачи пласта. Для исследования динамики коэффициента нефтеотдачи пласта были взяты 40 значений при промежуточной временной итерации для каждой пары выборки. Следовательно, общее количество набора данных составила $10086 \times 40 = 403440$.

Вязкость нефти варьируется в интервале 0.1 – 0.5, пористость в интервале 0.1 – 0.3 и различные варианты проницаемости.

Методы машинного обучения

В данной работе рассматривается задача обучение с учителем, которая является одним из классов задач в машинном обучении. Обучение с учителем подразумевает, что целевая функция y и признаки x измеряются, где главной целью является предсказания данных на основе обученных образцов. Наша задача относится к классу регрессионных задач с точки зрения методов машинного обучения. Синтетический набор данных, полученный из математической модели являются абсолютная проницаемость k , пористость p , вязкость μ , временная итерация t и коэффициент нефтеотдачи η . В

нашем случае, коэффициент нефтеотдачи представляется как целевая функция y , а остальные четыре данные представляются как признаки x .

$$x^{(i)} = \begin{bmatrix} k^{(i)} \\ p^{(i)} \\ \mu^{(i)} \\ t^{(i)} \end{bmatrix}, \quad i = 1, \dots, m \quad (1)$$

где $x^{(i)}$ является признаком на i^{th} тренировочном примере.

$$y^{(i)} = \eta^{(i)}, \quad i = 1, \dots, m \quad (2)$$

где $k^{(i)}$, $p^{(i)}$, $\mu^{(i)}$ и $t^{(i)}$ - абсолютная проницаемость, пористость, вязкость, временная итерация на i^{th} данных, и m является число тренировочных примеров (training example $m = 403440$). Таким образом, x является $(n_x + 1) \times m$ матрицей, а целевая функция y является $m \times 1$ вектором. Регрессионную модель можно записать в следующем виде:

$$y^{(i)} = h(x^{(i)}) + \varepsilon^{(i)}, \quad i = 1, \dots, m \quad (3)$$

где модель h описывает шаблон между x и y , и $\varepsilon^{(i)}$ является случайной ошибкой модели и измеряет некоторые расхождения. Для выполнения нашей задачи в данном исследовании был выбрана линейная регрессия (LR) и LR с полиномиальными свойствами.

Используя модель линейной регрессии, модель была обучена четырьмя входными параметрами и значениями коэффициента нефтеотдачи. В результате, обученная модель прогнозирует значения коэффициента нефтеотдачи на основе тестовых данных. Несмотря на то, что множественная линейная регрессия является очень простым, модель имеет несколько хороших преимуществ. Модель линейной регрессии освобождает инженера от необходимости хороших знаний в области физики в данном исследовании. Данная модель хорошо обучается и высококо интерпретируется, так как все независимые переменные множественной регрессии напрямую влияют к целевой функции. Следственно, влияние входных параметров легко обнаруживается и визуализируется.

В данной работе полиномиальная регрессия используется как частный случай множественной линейной регрессии. Так как, увеличение n степени полинома добавляет к линейной регрессии нелинейность данных. Однако это не означает, что с увеличением степени полинома модель будет обучаться еще лучше. Есть проблемы недобучения и переобучения модели. Для выбора оптимальной модели нужно находить компромисс между смещением и дисперсией.

Результаты и обсуждение

Данные были разделены на обучающий и тестовый набор данных. Для обучения был использован 8069 набор (80%) от общего объема данных, а для тестовой остальные 2017 пар (20%).

В качестве среды исполнения для машинного обучения был выбран язык программирования Python. Python является очень полезным для этой цели, так как имеет много библиотек, а также удобен в решении проблем связанными с машинным обучением. Для работы по импорту данных была использована библиотека Pandas. Визуализация данных был создан с помощью библиотеки Matplotlib. Используемые алгоритмы машинного обучения были выполнены с помощью библиотеки scikit-learn, которая предназначена для обучения данных с помощью многих методов машинного обучения. Как было упомянуто ранее, общее количество пар выборки составляет 10086 моделей. Каждый пар выборки состоит из 40 значения коэффициента нефтеотдачи. В результате, у нас есть много тестовых пар, однако в данной работе результаты будут показываться для некоторых тестовых пар. Рисунок 2 показывает результаты одной из тестовых пар выборки для разных методов регрессии.

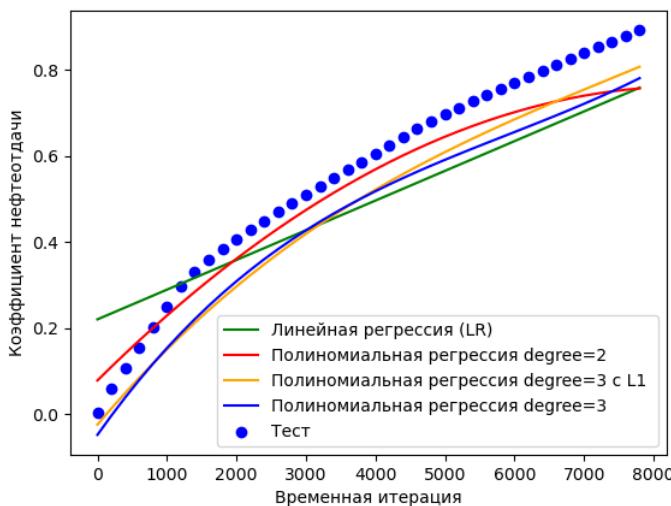


Рис. 2. Коэффициент нефтеотдачи на разных моделях регрессии

Мы можем видеть, что прогнозируемая функция линейной регрессии LR не захватывает все шаблоны в данных. Следственно, модель LR имеет пример недобучения (under-fitting).

С помощью полиномиальной регрессии (PR) увеличивается сложность модели. Для обучения с полиномиальными свойствами важно выбрать нужную модель, то есть степень полинома.

Из данного рисунка 2 совершенно понятно, что квадратичная полиномиальная модель обучает данные лучше, чем линейная модель. Также заметно, что кубическая полиномиальная модель прогнозирует данные хуже, чем квадратичная модель. Кубическая полиномиальная регрессия довольно хорошо прогнозирует целевую функцию для данной пары, однако из-за проблемы переобучения, другие пары из тестовой выборки могут плохо соответствовать данным. С помощью регуляризации L1 было улучшено кубическая полиномиальная модель для данной пары. В использовании L1 регуляризации было подобрано оптимальное значение λ . С помощью L1 регуляризации кубическая полиномиальная модель довольно хорошо прогнозирует функцию для всех тестовых данных, чем простая кубическая модель.

Следующая таблица демонстрирует среднюю оценку MSE для всех 20% тестовых наборов (Таблица 2).

Таблица 2. Оценка MSE для всех пар тестового набора

Алгоритмы машинного обучения	Тестовый набор (20%) MSE
Линейная регрессия	0.0037
Полиномиальная регрессия degree=2	0.0016
Полиномиальная регрессия degree=3	0.0084

Следующая таблица демонстрирует среднюю оценку R^2 для 80% обучающих наборов и 20% тестовых наборов (Таблица 3).

Таблица 3. Оценка R^2 для всех пар обучающего и тестового набора

Алгоритмы машинного обучения	Обучающий набор (80%) R^2	Тестовый набор (20%) R^2
Линейная регрессия	0.87	0.91
Полиномиальная регрессия degree=2	0.95	0.96
Полиномиальная регрессия degree=3	0.97	0.79
PR degree=3 с L1 регуляризации	0.96	0.92

Из данной таблицы заметно, что кубическая полиномиальная регрессия обучается при тренировочном наборе довольно хорошо, но на тестовых данных коэффициент детерминации R^2 уменьшается из-за высокой дисперсии между наборами данных при degree = 3.

Более того, можно заметить, что для тестовой выборки средняя квадратическая ошибка MSE увеличилась, а коэффициент детерминации R^2 уменьшился по сравнению с квадратичной моделью. Таким образом, для данной тестовой выборки самой оптимальной является квадратичная модель. Однако, это закономерность является правдивой только для данного набора. Для остальных пар из всей тестовой выборки, результаты могут быть другими. Это объясняется тем, что полиномиальная кубическая модель в нашем случае имеет переобучения из-за высокой дисперсии.

Заключение

Данная статья была посвящена к применению методов машинного обучения для прогнозирования добычи нефти. В данном исследовании в качестве метода машинного обучения была использована многомерная линейная регрессия с полиномиальными свойствами. Несмотря, что линейная регрессия является простым, данная модель хорошо обучается и высоко интерпретируется. Были протестированы разные степени полиномиальной регрессии, а так же было выявлено, что для наших синтетических данных квадратическая полиномиальная модель довольно хорошо обучается и отлично прогнозирует значение коэффициента нефтеотдачи. Для квадратичной модели полиномиальной регрессии коэффициент детерминации R^2 составляет 0.96, что является довольно хорошим результатом для тестовых данных. Таким образом, предполагается, что рассмотренные методы машинного обучения в данной статье могут быть полезными для прогнозирования добычи нефти на основе синтетических данных.

Признательность

Исследовательская работа выполнена при финансовой поддержке Комитета науки Министерства образования и науки Республики Казахстан (грант № BR05236447).

ЛИТЕРАТУРА

- [1] Krasnov, Fedor & Glavnov, Nikolay & Sitnikov, Alexander. (2018). A Machine Learning Approach to Enhanced Oil Recovery Prediction. 10.1007/978-3-319-73013-4_15.
- [2] Guo, Z., Reynolds, A. C., Zhao, H.: A Physics-Based Data-Driven Model for History-Matching, Prediction and Characterization of Waterflooding Performance. Society of Petroleum Engineers. doi:10.2118/182660-MS
- [3] Shehata, A. M., El-banbi, A. H., Sayyous, H.: Guidelines to Optimize CO₂ EOR in Heterogeneous Reservoirs. Society of Petroleum Engineers. doi:10.2118/151871-MS
- [4] Breiman, Leo: Random Forests Machine Learning 45 (1): 532 DOI:10.1023/A:1010933404324
- [5] Aliyuda, Kachalla & Howell, John. (2019). Machine Learning Algorithm for Estimating Oil Recovery Factor Using a Combination of Engineering and Stratigraphic Dependent Parameters. Interpretation. 7. 1-34. 10.1190/int-2018-0211.1.
- [6] Mirzaei-Paiaman, Abouzar & Salavati, Soroor. (2012). The Application of Artificial Neural Networks for the Prediction of Oil Production Flow Rate. Energy Sources. 34. 10.1080/15567036.2010.492386.
- [7] Jreou, Ghazwan. (2012). : Application of neural network to optimize oil field production. Asian Transactions on Engineering. 3. 1-9.
- [8] Ristanto, Tita & Horne, Roland. (2018). Machine Learning Applied to Multiphase Production Problems.
- [9] Horne, R. N. 2007. Listening to the Reservoir—Interpreting Data From Permanent Downhole Gauges. J Pet Technol 59 (12): 78–86. SPE-103513-JPT. <https://doi.org/10.2118/103513-JPT>.
- [10] Liu, Y. and Horne, R. N. 2011. Interpreting Pressure and Flow Rate Data From Permanent Downhole Gauges Using Data-Mining Approaches. Presented at the SPE Annual Technical Conference and Exhibition, Denver, 30 October–2 November. SPE-147298-MS. <https://doi.org/10.2118/147298-MS>.
- [11] Liu, Y. and Horne, R. N. 2012. Interpreting Pressure and Flow-Rate Data From Permanent Downhole Gauges by Use of Data-Mining Approaches. SPE J. 18 (1): 69–82. SPE-165346-PA. <https://doi.org/10.2118/165346-PA>.
- [12] Liu, Y. and Horne, R. N. 2013. Interpreting Pressure and Flow Rate Data From Permanent Downhole Gauges Using Convolution-Kernel-Based DataMining Approaches. Presented at the SPE Western Regional & AAPG Pacific Section Meeting 2013 Joint Technical Conference, Monterey, California, 19–25 April. SPE-165346-MS. <https://doi.org/10.2118/165346-MS>.
- [13] Liu, Y. and Horne, R. N. 2013. Interpreting Pressure and Flow Rate Data From Permanent Downhole Gauges With Convolution-Kernel-Based DataMining Approaches. Presented at the SPE Annual Technical Conference and Exhibition, New Orleans, 30 September–2 October. SPE-166440-MS.<https://doi.org/10.2118/166440-MS>.
- [14] Tian, Chuan & Horne, Roland. (2019). Applying Machine-Learning Techniques To Interpret Flow-Rate, Pressure, and Temperature Data From Permanent Downhole Gauges. SPE Reservoir Evaluation & Engineering. 22. 386-401. 10.2118/174034-PA.

[15] Cao, Q. & Banerjee, R. & Gupta, S. & Li, J. & Zhou, W. & Jeyachandra, B.. (2016). Data Driven Production Forecasting Using Machine Learning. 10.2118/180984-MS.

[16] Shengnan, C. 2019. Application of Machine Learning Methods to Predict Well Productivity in Montney and Duvernay. Calgary Petroleum Club.

Кенжебек Е.Ғ., Иманкулов Т.С., Ахмед-Заки Д.Ж.

Полиномды регрессия әдісімен мұнай өндіруді болжау

Түйінде. Бұл мақалада машиналық оқыту әдістерін қолдана отырып, мұнай өндіруді болжау жұмысы ұсынылған. Машиналық оқыту әдісі ретінде көп өлшемді сызықтық регрессия алгоритмі полиномдық қасиеттермен жүзеге асырылды. Синтетикалық мәліметтер жиынтығы гидродинамика есептеу және мұнай өндіру есептеріндегі қанықтылықтың таралуын анықтау үшін қолданылатын Баклей-Леверетт математикалық моделінің көмегімен алынды. Мұнай өндіру мәселесінің параметрлерінің әртүрлі комбинациясы таңдалды, онда кеуектілік, мұнай фазасының тұтқырығы және жыныстың абсолютті еткізгіштігі машиналық оқытуға арналған кіріс параметрлері ретінде алынды. Ал шығу параметрі ретінде мұнай беру коэффициентінің мәні таңдалды. Полиномдық регрессия алгоритмін тестілеу үшін 10086 жұп деректер пайдаланылды. Машиналық оқыту үшін орындау ортасы ретінде Python бағдарламаласа тілі қолданылды. Полиномдық регрессияның әртүрлі дәрежелері жүзеге асырылды және синалды, сонымен қатар біздің синтетикалық деректеріміз үшін квадраттық полиномдық модель өте жақсы оқытылатыны және мұнай беру коэффициентінің мәнін жақсы болжайтыны аныкталды. Қайта оқыту мәселесін шешу үшін Лассо регрессия әдісі деп аталатын L1 регуляризаторы қолданылды. Полиномдық регрессияның квадраттық моделі үшін R^2 анықтау коэффициенті 0.96 құрады, бұл синақ деректері үшін өте жақсы нәтиже.

Кілт сөздер: мұнай өндіруді арттыру, машиналық оқыту, регрессия әдісі, полиномдық регрессия, регуляризация.

UDK 004.9

¹CH. Kenshimov, ^{2,3}Iu. Krak, ¹Zh.Amirgalieva, ²S. Kondratuk, ⁴A. Aitimov

(¹Institute of Information and Computational Technologies, Almaty, Kazakhstan

²Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

³V.M.Glushkov Institute of Cybernetics, Kyiv, Ukraine

⁴Suleyman Demirel University, Kaskelen, Kazakhstan, e-mail: amir_ed@mail.ru)

INFORMATION TECHNOLOGY FOR SIGN LANGUAGE ALPHABET UNITS MODELING AND RECOGNITION

Abstract. The information technology, which is implemented with cross platform tools, is proposed for modeling and recognition of gesture units of sign language, animation between states of gesture units with a combination of gestures (words). Implemented technology simulates sequence of gestures using virtual spatial hand model and performs recognition of dactyl items from camera input. Convolutional neural networks are used for dactyl items recognition. With the cross platform means technology achieves the ability to run on multiple platforms without re-implementing for each platform. Information and gesture communication technology was developed with further scaling capabilities in mind for gestures of other languages alphabets.

Key words: information technology, cross platform, sing language, fingerspelling alphabet, convolutional neural networks.

¹Ш. Кеншимов, ^{2,3}Ю. Крак, ¹Ж. Амирғалиева, ²С. Кондратюк, ^{1,4}А.Айтимов

(¹Институт информационных и вычислительных технологий КН МОН РК, Алматы, Казахстан

²Киевский национальный университет им. Т. Шевченко, Киев, Украина

³Институт кибернетики им. В.М.Глушкова, Киев, Украина

⁴Университет Сулеймана Демиреля, Каскелен, Казахстан

e-mail: amir_ed@mail.ru)

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ МОДЕЛИРОВАНИЯ И РАСПОЗНАВАНИЯ ЗНАКОВ ЖЕСТОВОГО ЯЗЫКА

Аннотация. Информационная технология, реализованная с использованием кроссплатформенных инструментов, предлагается для моделирования и распознавания жестовых единиц языка жестов, анимации между состояниями жестовых единиц с помощью комбинации жестов (слов). Внедренная технология имитирует последовательность жестов с использованием виртуальной пространственной модели руки и выполняет