УДК 681.518:622.276
DOI 10.25205/1818-7900-2020-18-1-27-35

Об одном методе классификации нефтяного месторождения с использованием комплекса геолого-промысловых данных и машинного обучения

Д. В. Курганов

*Самарский государственный технический университет
Самара, Россия*

Аннотация

Применение алгоритмов машинного обучения в частности для классификации и зонирования месторождений нефти и газа, является перспективным направлением при анализе разработки месторождений. Существенным условием применения метода является наличие обширной цифровой базы с представительными результатами. В работе рассмотрено применение метода кластеризации k -средних для крупного месторождения в Западной Сибири, а также проведен анализ эффективности применяемых гидроразрывов пласта. Методика позволяет комплексировать такие параметры, как плотность текущих запасов, географическое расположение скважин, параметры проведенных гидроразрывов пласта, пористость, нефтенасыщенность. В целом это позволяет учесть максимум геолого-промысловой информации, создавая предпосылки к автоматизации управления разработкой нефтяного месторождения. Также существенно увеличивается точность прогноза продуктивности коллектора, особенно имеющего сложное строение, в том числе техногенно модифицированного.

Ключевые слова

машинное обучение, классификация, дебит, метод k -средних, нефть, скважина, выборка, месторождение, запасы, гидроразрыв пласта, продуктивность, диаграмма Вороного

Для цитирования

Курганов Д. В. Об одном методе классификации нефтяного месторождения с использованием комплекса геолого-промысловых данных машинного обучения // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 27–35. DOI 10.25205/1818-7900-2020-18-1-27-35

Oil Reservoir Classification by Geological and Production Data Using Unsupervised Machine Learning Algorithm

D. V. Kurganov

*Samara State Technical University
Samara, Russian Federation*

Annotation

In machine learning, k -means unsupervised model is used for classification analysis. In this paper k -means model is applied for productivity prediction of giant Western Siberian oilfield. An essential condition for method's application is availability of digital databases with representative results. Complex method allows combine different reservoir and production parameters: rates, porosity, saturation, frac parameters etc. The method can be particularly useful in complicated reservoirs, e.g. in dual porosity ones, where the relationship between formation parameters (permeability, porosity, saturation) and production rates is unclear and cannot be set by traditional development analysis, particularly in frac environment.

Keywords

machine learning, classification, rate, k -means, oil, well, reservoir, oil in place, sample, frac, productivity, Voronoi diagram

© Д. В. Курганов, 2020

For citation

Kurganov D. V. Oil Reservoir Classification by Geological and Production Data Using Unsupervised Machine Learning Algorithm. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 27–35. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-27-35

Машинное обучение (МО) в настоящее время является перспективным методом, получившим стремительное развитие в нефтегазовой отрасли [1].

Наряду с накоплением фактического и исторического материала по эксплуатации скважин и месторождений появляются новые возможности для установления различных закономерностей и обобщений, связанных с конкретными залежами или типами залежей. Комплекс геолого-промысловых данных, имеющийся на месторождении, подразумевает проведение определенного анализа известными методами МО [2]. В частности, имеется возможность применить к ним известные алгоритмы классификации для выявления зон с совокупностью определенных характеристик, непосредственно влияющих на продуктивность скважин [3]. В отличие от стандартного картирования свойств и дальнейшего анализа набора карт, применяемого в отрасли, данные методы позволяют учесть сразу всю совокупность признаков, влияющих на потенциальную добычу углеводородов, что позволяет значительно упростить процедуру размещения проектных скважин для качественной выработки запасов.

Кластеризация методом k -средних [4] является одним из так называемых неконтролируемых методов МО, в котором исходные данные разбиваются на группы по совокупности определенных признаков.

Процедура кластеризации выглядит следующим образом.

- Первоначально центры кластеров выбираются случайно.
- От каждой точки в пространстве параметров рассчитывается Евклидово расстояние до центра ближайшего кластера.
- Итеративно минимизируется сумма расстояний от каждой точки до центра соответствующего кластера.

Формально задача ставится следующим образом. Пусть имеется m наблюдений (из пространства R^n). Необходимо разбить эти наблюдения на k кластеров, при этом каждое наблюдение относится к тому кластеру, к центру (так называемому центроиду) которого оно ближе всего.

В качестве меры близости используется Евклидово расстояние:

$$\rho(x, y) = \|x - y\| = \sqrt{\sum_{p=1}^n (x_p - y_p)^2},$$

где $x, y \in R^n$.

Таким образом, для ряда наблюдений $(x^{(1)}, x^{(2)}, \dots, x^{(m)})$, $x^j \in R^n$ метод k -средних разделяет m наблюдений на k групп (или кластеров) ($k \leq m$) $S = \{S_1, S_2, \dots, S_k\}$, таким образом, чтобы минимизировать суммарное квадратичное отклонение точек кластеров от центроидов этих кластеров:

$$\min \left[\sum_{i=1}^k \sum_{x^{(j)} \in S_i} \|x^{(j)} - \mu_i\|^2 \right],$$

где $x^{(j)} \in R^n$, $\mu_i \in R^n$, μ_i – центроид для кластера S_i .

Если мера близости до центроида определена, то разбиение объектов на кластеры сводится к определению центроидов этих кластеров. Число кластеров k задается заранее.

Рассмотрим первоначальный набор k -средних (центроидов) $\mu_1, \dots, \mu_k$ в кластерах $S_1, \dots, S_k$. Первоначально центроиды кластеров выбираются случайно или по определенному правилу (например, можно выбрать центроиды, максимизирующие начальные расстояния между кластерами).

Далее следует отнести наблюдения к тем кластерам, чье среднее (центроид) к ним ближе всего. Каждое наблюдение принадлежит только к одному кластеру, даже если его можно отнести к двум и более кластерам.

Затем центроид каждого i -го кластера перевычисляется по следующему правилу:

$$\mu_j = \frac{1}{S_j} \sum_{x^{(j)} \in S_j} x^{(j)}.$$

Таким образом, алгоритм k -средних заключается в перевычислении на каждом шаге центроида для каждого кластера, полученного на предыдущем шаге. Алгоритм останавливается, когда значения μ_i перестают меняться.

Основной проблемой данного алгоритма является то, что оптимальное значение кластеров заранее неизвестно. Одной из методик для его нахождения является определение так называемого *силуэта* выборки [5]. Данный коэффициент не предполагает знания истинных меток объектов и позволяет оценить качество кластеризации, используя только саму неразмеченную выборку и результат кластеризации.

Пусть a – среднее расстояние от данного объекта до объектов из того же кластера, b – среднее расстояние от данного объекта до объектов из ближайшего кластера (отличного от того, в котором лежит сам объект). Тогда силуэтом данного объекта называется величина

$$s = (b - a) / \max(a, b).$$

Силуэтом выборки называется средняя величина силуэта объектов данной выборки. Таким образом, силуэт показывает, насколько среднее расстояние до объектов своего кластера отличается от среднего расстояния до объектов других кластеров. Данная величина лежит в диапазоне $[-1, 1]$. Значения, близкие к -1 , соответствуют плохим (разрозненным) кластеризациям; значения, близкие к нулю, говорят о том, что кластеры пересекаются и накладываются друг на друга; значения, близкие к 1 , соответствуют «плотным» четко выделенным кластерам. Таким образом, чем больше силуэт, тем более четко выделены кластеры, и они представляют собой компактные, плотно сгруппированные облака точек.

Также для дальнейшего изложения необходимо привести определение диаграммы Вороного [6]. Диаграмма Вороного P_i , построенная для точки измерений x_i , характеризуется тем, что содержит те и только те точки, расстояние от которых до точки x_i меньше или равно расстоянию до любой другой точки измерений x_j . При построении диаграмм Вороного используется система соседства, полученная в процессе триангуляции Делоне. Границы диаграмм Вороного P_i состоят из отрезков серединных перпендикуляров, проведенных к сторонам треугольников Делоне.

Применительно к разработке месторождений границы диаграммы Вороного могут рассматриваться как границы зон дренирования скважины.

В качестве примера взято одно из месторождений Западной Сибири, пласт БВ₁₃. Залежь пластово-сводовая, строение сложное. Пласт характеризуется невысокой проницаемостью, значительной изменчивостью параметров по площади. На объекте пробурено 180 скважин (87 действующих добывающих, 50 действующих нагнетательных), реализована очаговая система заводнения, всего было проведено около 90 гидроразрывов пласта (ГРП). Текущая степень выработки 67 %, средняя обводненность продукции – 88 %. Средний дебит скважин по жидкости составляет 60 т/сут. Нефть маловязкая, с высоким газосодержанием. Стандартные методы анализа процесса разработки, в частности фильтрационное моделирование, не

позволяют детально раскрыть процессы, происходящие в пластовой системе. Адаптацию модели нельзя признать удовлетворительной – по-видимому, в процессе проведения многочисленных ГРП в пластовой системе образовалась вторая среда (система трещин), течение флюидов в которой не описывается стандартными законами фильтрации [7].

Для анализа и классификации были выбраны следующие данные: интерпретированный каротаж (пористость, насыщенность, эффективная толщина), добыча первого года промышленной эксплуатации каждой скважины, информация об использованном при ГРП количестве проппанта и разрывающей жидкости. Это важнейшие количественные характеристики процесса ГРП, которые косвенно позволяют судить о его эффективности по данной скважине. В выборку попали скважины, по которым был проведен ГРП после начала закачки на месторождении.

В дополнение к скважинным данным была рассчитана плотность текущих запасов по каждой диаграмме Вороного. Текущие запасы были рассчитаны на основе начальных запасов за вычетом накопленных отборов по каждой скважине [8].

Коллектор был зонально разделен на три типа посредством кластеризации (Класс А, Класс Б и Класс С). С помощью расчета силуэта было установлено, что добавление четвертого класса не оказывает существенного влияния на понимание характеристик пласта и характер разработки месторождения.

В качестве параметров для кластеризации были выбраны следующие: количество проппанта, количество жидкости разрыва, площадное месторасположение скважин, плотность текущих удельных запасов на скважину по диаграмме Вороного, а также величина, характеризующая изначальное качество коллектора в смысле потенциального содержания запасов – произведение начальной нефтенасыщенности, пористости и эффективной толщины ($S_0 \cdot \phi \cdot H_{\text{eff}}$). На рис. 1 приведен результат работы алгоритма с выделением скважин по описанной совокупности признаков. Рисунок 2 показывает примерное эмпирическое зонирование месторождения на основе разбиения по классам скважин, рис. 3 иллюстрирует распределение признаков по классам. Средние значения признаков приведены на рис. 4. Следует отметить, что все исходные данные прошли процедуру нормализации (приведены к интервалу [0, 1] для устранения эффекта масштабирования. Северная зона (розовый цвет) характеризуется худшими показателями, центральная зона – удовлетворительными, южная – лучшими показателями (см. рис. 2). Все распределения параметров, так же как и средние значения, улучшаются от северной части к южной. Особенно важным представляется тот факт, что лучшая зона характеризуется большими трещинами (а точнее, большим количеством проппанта и разрывающей жидкости). Добывающие скважины в коллекторе лучшего качества и с большими трещинами в среднем характеризуются большей добычей.

При этом следует отметить, что в процессе кластеризации начальные дебиты нефти в качестве параметров не использовались, чтобы более точно выявить влияние всех остальных параметров пластовой системы на фактическую добычу.

В среднем добыча нефти скважин класса В превышает добычу скважин класса А в 1,5 раза, а добыча скважин класса С превышает добычу скважин класса В в 1,3 раза (рис. 5).

Чтобы исследовать влияние ГРП на добычу скважин, необходимо выбрать лишь те скважины, которые находятся в качественном коллекторе, исключив из анализа остальные скважины, а именно те, у которых значение ($S_0 \cdot \phi \cdot H_{\text{eff}}$) меньше среднего по классу С. При этом средним значением является 0,7. На рис. 6 цветом выделены скважины, попавшие в качественный коллектор.

Рисунок 7 иллюстрирует влияние параметров проведенных гидроразрывов на отборы нефти. Видно, что наличие качественного коллектора не гарантировало больших дебитов. В то же время большие объемы проппанта и разрывающей жидкости вели в среднем к увеличенным отборам.

Рис. 1. Фрагмент месторождения с выделенными скважинами по методу k -средних (красные – зона А, синие – зона В, зеленые – зона С)

Fig. 1. Part of the oilfield with colored wells ranked by k -means algorithm (red – zone A, blue – zone B, green – zone C)

Рис. 2. Фрагмент месторождения с выделенными зонами по методу k -средних

Fig. 2. Part of the oilfield with zones ranked by k -means algorithm

Рис. 3. Распределения нормализованных параметров модели кластеризации

Fig. 3. Distribution of cluster model parameters

Рис. 4. Средние значения параметров по классам коллектора (нормированные)

Fig. 4. Average parameters distribution classified by reservoir type

Рис. 5. Распределение дебитов по классам коллектора

Fig. 5. Oil rates classified by reservoir type

Рис. 6. Распределение скважин, попавших в качественный коллектор

Fig. 6. Well locations of supreme quality reservoir

Рис. 7. Распределение параметров модели по скважинам качественного коллектора
 Fig. 7. Average parameters distribution of supreme quality reservoir

Выполненная классификация позволяет оценить различные зоны месторождения на предмет его дальнейшего разбуривания и потенциальной продуктивности с учетом проводимых гидроразрывов пласта, позволяет выявлять значимые параметры и объемы проведения геолого-технических мероприятий, влияющих на продуктивность скважин. Также описанная методика и найденные классы могут использоваться, например, для контролируемого машинного обучения при создании моделей уплотняющего бурения. Обучив модель на найденных образцах, можно прогнозировать продуктивность новых скважин и зон месторождения, в том числе осложненных средой с двойной пористостью [9], принимать управляющие решения для дальнейшей разработки.

Список литературы

1. **Zangl G., Hannerer J.** Data Mining: Applications in the Petroleum Industry. Katy, TX, Round Oak Publ., 2003, 222 p.
2. **Hand D., Manilla H., Smyth P.** Principles of Data Mining. MIT Press, 2001, 546 p.
3. **Курганов Д. В.** О численном решении одной задачи минимизации в моделировании пластовых систем // Сибирский журнал индустриальной математики. 2003. Т. 6, № 1. С. 51–59.
4. **Han J., Kamber M., Pei J.** Data Mining: Concepts and Techniques. USA, Elsevier, 2012, 703 p.
5. **Wu X., Kumar V.** The Top Ten Algorithms in Data Mining. USA, Taylor & Francis Group, 2009, 201 p.

6. Демьянов В. В., Савельева Е. А. Геостатистика: теория и практика. М.: Наука, 2010. 327 с.
7. Курганов Д. В., Попков В. И., Хмелевских Е. И. Оценка влияния низкопроницаемых прослоев коллектора на эффективность выработки запасов нефти с использованием трехмерных гидродинамических моделей // Изв. Самарского Научного центра РАН. 2002. Спец. выпуск: Проблемы нефти и газа. С. 121–126.
8. Колесов В. В., Курганов Д. В. Интегрированное моделирование неоднородности и связности гигантского карбонатного резервуара по геофизическим и промысловым данным на поздней стадии разработки с целью локализации остаточных запасов нефти // Материалы технической конференции SPE «Разработка зрелых месторождений». М., 2017. С. 121–129.
9. Смехов Е. М., Дорофеева Т. В. Вторичная пористость горных пород – коллекторов нефти и газа. Л.: Недра, 1987. 96 с.

References

1. Zangl G., Hannerer J. Data Mining: Applications in the Petroleum Industry. Katy, TX, Round Oak Publ., 2003, 222 p.
2. Hand D., Manilla H., Smyth P. Principles of Data Mining. MIT Press, 2001, 546 p.
3. Kurganov D. V. Numerical solution of one minimization problem in reservoir simulation. *Sibirskiy zhurnal industrial'noy matematiki*, 2003, vol. 6, no. 1, p. 51–59. (in Russ.)
4. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. USA, Elsevier, 2012, 703 p.
5. Wu X., Kumar V. The Top Ten Algorithms in Data Mining. USA, Taylor & Francis Group, 2009, 201 p.
6. Demiyarov V. V., Savelyeva E. A. Geostatistika: teoriya i praktika. Moscow, Nauka, 2010, 327 p. (in Russ.)
7. Kurganov D. V., Popkov V. I., Khmelevskikh E. I. Estimation of low permeability layers on field development process using reservoir simulation. *Izvestiya Samarskogo Nauchnogo tsentra RAN*, 2002, Special issue, p. 121–126. (in Russ.)
8. Kolesov V. V., Kurganov D. V. Integrated modeling of heterogeneity and conductivity of giant mature carbonate reservoir using complex geophysic and production data. In: Proc. of SPE Conference “Razrabotka zrelykh mestorozhdeniy”. Moscow, 2017, p. 121–129. (in Russ.)
9. Smekhov E. M., Dorofeeva T. V. Dual porosity oil and gas reservoirs. Leningrad, Nedra Publ., 1987, 96 p. (in Russ.)

Материал поступил в редколлегию

Received
04.09.2019

Сведения об авторе

Курганов Дмитрий Владимирович, кандидат физико-математических наук, доцент, доцент кафедры разработки и эксплуатации нефтяных и газовых месторождений, Самарский государственный технический университет (Самара, Россия)
Dmitri.Kourganov@inbox.ru

Information about the Author

Dmitri V. Kurganov, PhD, Smara State Technical University (Samara, Russian Federation)
Dmitri.Kourganov@inbox.ru