

РАЗРАБОТКА МЕТОДОВ КЛАССИФИКАЦИИ ТИПОВ ПОРОД И АВТОМАТИЧЕСКОГО ВЫДЕЛЕНИЯ ЛИТОТИПОВ ДЛЯ КАРБОНАТНЫХ КОЛЛЕКТОРОВ ПО СОВОКУПНОСТИ ОПРЕДЕЛЯЮЩИХ ПАРАМЕТРОВ

И.Р. Миннихметов, А.Х. Пергамент
ИПМ им. М.В.Келдыша РАН, e-mail: office@keldysh.ru

Введение

На сегодняшний день при освоении месторождений углеводородов и решении задач оптимизации разработки широко применяются методы математического моделирования. Основой такого подхода является использование постоянно действующих геолого-гидродинамических технологических цифровых моделей месторождений (ПДГТМ).

Одним из важнейших этапов построения ПДГТМ является литолого-фациальное моделирование и определение областей распространения пород-коллекторов на основе доступных данных наблюдений. Наиболее точные результаты могут быть получены с помощью отбора и последующего исследования образцов керна из потенциально продуктивных отложений. Такие исследования являются сложными и дорогостоящими и зачастую не проводятся массово, в отличие от геофизических исследований скважин (ГИС).

Использование результатов ГИС позволяет достоверно классифицировать типы пород, составляющие терригенную толщу [1, 2]. В случае же карбонатных коллекторов такую задачу трудно решить с использованием методов ГИС. Поэтому для карбонатных коллекторов разработка методики автоматического выделения типов пород при ограниченном количестве кернового материала имеет большое научное и практическое значение.

В настоящей работе предлагается использовать методы машинного обучения при построении эффективных классифицирующих алгоритмов литофациального анализа, в частности, машины опорных векторов. Для снижения размерности пространства исходных данных используются способы снижения размерности на основе нелинейного метода главных компонент. Особенностью работы является вероятностный подход, позволяющий определить степень достоверности полученных результатов.

1. Постановка задачи

Пусть задана система n классов $C = \{c_1, c_2, \dots, c_n\}$, элементами которой являются различные карбонатные породы, например: известняки тонко-микрозернистые с криноидно-мшанковым детритом (вакстоуны), известняки органогенно-обломочные криноидно-мшанковые с тонко-микрозернистым цементом (пакстоуны) и др. Также задано пространство параметров $X = \{X^g \oplus X^k\}$, элементами которого служат данные ГИС X^g и керны X^k . Данные ГИС представлены набором каротажных кривых: БК, АК, НК и др. Данные керновых исследований – пористость и проницаемость по керну, смачиваемость, цветность керна и др. Проинтерпретированные данные, то есть множество пар $\{x_i^0 \in X, c_i^0 \in C\}, i = 1..M$, задают набор обучающих данных. На основе обучающих данных необходимо построить отображение $F: X \rightarrow C$ пространства наблюдаемых параметров X во множество типов $C = \{c_1, c_2, \dots, c_n\}$.

Ввиду того, что количество кернового материала сильно ограничено, задача выделения литотипов решается в два этапа:

1) на первом этапе требуется построить вероятностный классификатор на основе обучающего набора данных, позволяющий определить распределение классов $P[F(x) = c_i], i = 1..n$ для произвольной точки x пространства $X = \{X^g \oplus X^k\}$. Данная задача является классической задачей классификации с "учителем" [3]. Для наглядности на рис. 1 представлен пример задачи классификации в двумерном пространстве параметров. Круги соответствуют экспериментальным данным, на основе которых проводится разделение всего пространства параметров на классы. Классы отмечены розовым, голубым и светло-зеленым фоном. Отметим, что на данном этапе, для выделения литотипов используются как данные ГИС, так и результаты керновых исследований.

2) на втором этапе на основе построенного вероятностного классификатора требуется определить вероятность $P[F(x^g) = c_i], i = 1..n$ для произвольной точки x^g пространства X^g . Таким образом, построенный ранее классификатор применяется для определения литотипа только с использованием данных ГИС. На рис. 1 продемонстрирован случай двумерного пространства параметров, где один параметр соответствует кривой ГИС, а другой – данным исследования керна. Как мы видим,

вероятность принадлежности точки x^* к классам, представленным красными, синими и зелеными точками, пропорционально длине отрезков, на которые делится пунктирная линия. Для изображенного случая, наиболее вероятным является класс, обозначенный голубым цветом.

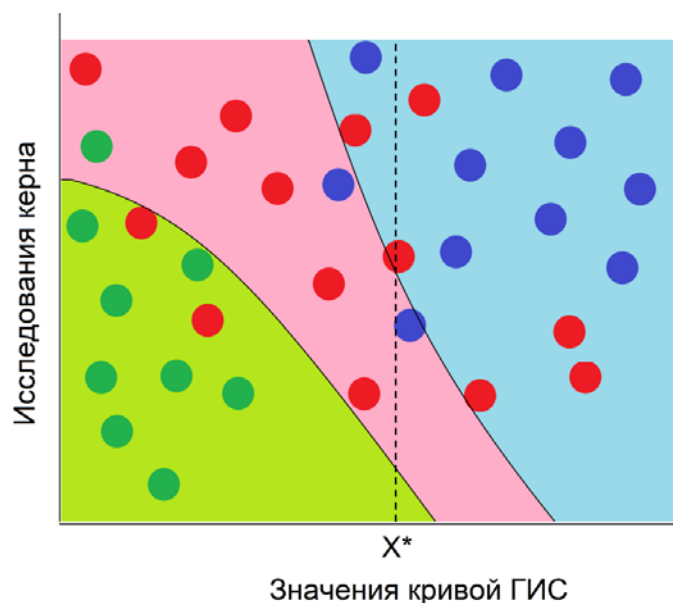


Рис. 1. Задача классификации в двумерном пространстве параметров X . Кружки соответствуют экспериментальным данным, на основе которых проводится разделение всего пространства параметров на классы (розовая, голубая и светло-зеленая заливка)

2. Характеристики литотипов

В данной работе задача построения отображения $F: X \rightarrow C$ сводится к построению классификатора для определения характерных признаков литотипов и дальнейшего восстановления литотипа по полученным характеристикам.

В качестве характеристик пород использовались разработки профессора А.В. Постникова. Согласно предложенной им классификации, карбонатные породы можно условно разделить: по «подклассу», «структуре», «биоте», «спаритизации», «выщелачиванию», «инкорпорации», «вторичному минерало-образованию», «типу коллектора», «емкости».

К примеру, известняки биогермные сферово-водорослевые с граноморфным наполнителем пористые (баундстоуны) имеют следующие характеристики: подкласс: «известняки смешанные, комковато-сгустковые», структура: «оолитовые», спаритизация «слабая», выщелачивание: «слабое», инкорпорация «сильная», доломитизация: «слабая», тип коллектора: «порово-каверновый первичный», емкость 20-30%.

3. Анализ главных компонент

Задача построения искомого классификатора является задачей аппроксимации функции многих переменных. Как известно [3–5], ввиду большого пространства возникают большие трудности. С целью уменьшения количества входных переменных использовался нелинейный иерархический метод анализа главных компонент (h-NLPCA), предложенный [6]. Согласно результатам, полученным в работах [6, 7], данный подход позволяет эффективно выделить главные компоненты для нелинейных случаев.

4. Классифицирующий алгоритм

Для построения функций $f_k(x)$ использовался алгоритм машины опорных векторов (МОВ). Как известно [8, 9], МОВ является широко используемым бинарным классификатором. В простейшем случае МОВ позволяет определить гиперплоскость в пространстве параметров X , разделяющих классы признаков, причем, граница разделения классов признаков будет максимальной. В этом случае определяемая МОВ гиперплоскость называется оптимальной (рис. 2).

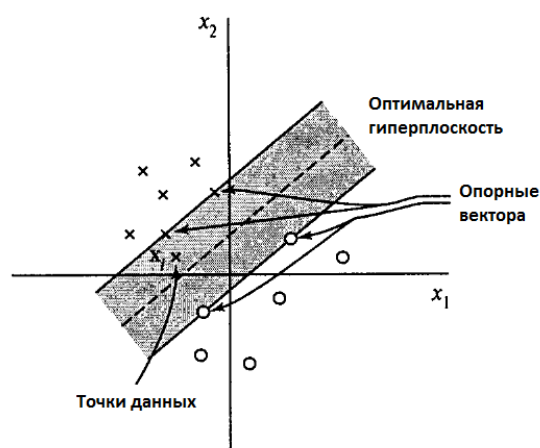


Рис. 2. Опорные вектора для решения задачи линейной бинарной классификации

Классические машины опорных векторов решают задачу линейной бинарной классификации. Для применения нелинейного мультиклассового подхода использовались данные работ [10, 11]. В соответствии с ними строятся отдельные нелинейные МОВ для каждого класса. Таким образом, каждая МОВ является индикатором принадлежности конкретной точки пространства к данному классу. Искомый класс для произвольной

точки пространства X , определяется с помощью вероятностного подхода: предполагается, что вероятность принадлежности x к классу c_i :

$$P(c_i | \vec{x}) = \frac{1}{1 + \exp(Ag(\vec{x}) + B)}, \quad (1)$$

где функция $g(\vec{x})$ – выходной сигнал машины опорных векторов, а константы A и B определяются методом максимального правдоподобия для тренировочного набора.

В конечном итоге мы получаем вероятности принадлежности каждой точки пространства X к конкретному классу. Далее для любой точки пространства X мы можем

определить класс, разыгрывая случайную величину, согласно построенной функции вероятности или выбирая максимально вероятный класс.

5. Классификация на основе данных ГИС

Следующий этап работы заключается в определении вероятности $P(c_i | x^g)$ на основе только данных ГИС x^g и построенной ранее вероятности $P(c_i | x)$. Согласно теореме Байеса, вероятность $P(c_i | x^g)$ определяется формулой:

$$P(c_i | x^g) = \int_{\tau \in X^k} P(c_i, x^k = \tau | x^g) \rho_{x^k}(\tau) d\tau, \quad (2)$$

где $\rho_{x^k}(\tau)$ – плотность распределения x^k , вероятность $P(c_i, x^k = \tau | x^g)$ принадлежности точки $x = \{x^k = \tau, x^g\}$ классу c_i определяется ранее построенным вероятностным классификатором $P(c_i | x)$.

Для решения подобных задач используется метод Монте–Карло [12]. Интегрирование методом Монте–Карло основано на вычислении интеграла по случайной выборке точек $\{x_1^k, x_2^k, \dots, x_q^k\}$ из пространства параметров X^k с заданной плотностью вероятности $p(x^k)$. Тогда оценка интеграла (2) имеет вид:

$$\hat{\mu}(c_i, x^g) = \sum_{i=1}^q \frac{P(c_i, x_i^k | x^g)}{p(x_i^k)}. \quad (3)$$

Ввиду того, что функция $P(c_i, \tau | x^g)$ является многомерной и имеет очень сложный вид, использование равномерного распределения x^k требует огромного числа элементов выборки. Для улучшения сходимости оценки (3) был разработан алгоритм, позволяющий разыгрывать значения $\{x_1^k, x_2^k, \dots, x_q^k\}$ с функцией распределения, максимально близкой к распределению $P(c_i, \tau | x^g)$.

Зная вероятности принадлежности к классу для произвольной точки пространства X , вычисляется наиболее вероятный класс и достоверность результата.

6. Результаты

В настоящем разделе рассмотрен пример применения разработанного подхода для реальных данных.

Входными данными для алгоритма являются:

- 1) скважинные кривые: гамма-каротаж, акустический каротаж, пористость, проницаемость, нефтенасыщенность.
- 2) петрофизические исследования керна, минералогия и данные смачиваемости по керну.
- 3) результаты оптического исследования керна в дневном и ультрафиолетовом диапазоне.
- 4) литотипы, выделенные экспертом.

Параметры модели:

– 100 скважин, при этом для 8 скважин известны все исходные данные (в частности, известны литотипы);

– 8 литотипов: известняки криноидно-мшанковые (пакстоуны), известняки криноидно-мшанковые (грейнстоуны), известняки фораминиферово-водорослевые (грейнстоуны), известняки биогермные палеоплизиновые, известняки биогермные микрокодиевые, известняки биогермные водорослевые, известняк биогермный тубифитовый.

Весь массив информации был разделен на две группы: данные для обучения машины опорных векторов и тестовые данные, на которых проводилась валидация алгоритма.

На первом этапе пространство параметров было разделено на классы характеристик. В качестве входных данных использовались данные ГИС и керна.

На основе выделенных характеристик были определены литотипы и сопоставлены с тестовыми данными. Ошибка алгоритма составила менее 7%. На рис. 3 продемонстрированы результаты применения методики для одной из тестовых скважин.

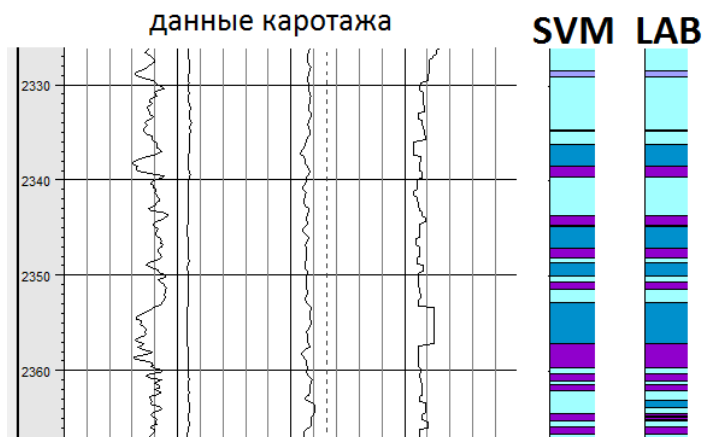


Рис. 3. Сравнение работы алгоритма и результатов лабораторных исследований (слева – кривая ГИС, в центре – результат автоматической классификации, справа – результат лабораторных исследований керна)

На втором этапе построенный вероятностный классификатор использовался для выделения характеристик только на основе данных ГИС методом Монте–Карло.

Далее по полученным характеристикам были предопределены литотипы и сопоставлены с лабораторными данными. Общая ошибка метода при этом составила менее 10%.

Согласно рис. 4, точки, в которых зафиксировано отклонение прогноза от факта, имеют низкую степень достоверности прогноза. Таким образом, данный метод позволяет не только предопределить наиболее вероятный литотип, но и оценить степень неопределенности результата.

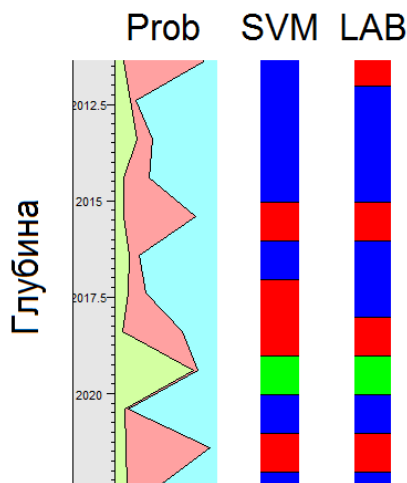


Рис. 4. Сравнение работы алгоритма и результатов лабораторных исследований (столбец слева соответствует значениям вероятности, в центре – результат автоматической классификации, справа – результат лабораторных исследований керна)

Выводы

Особенностью предложенной методики автоматического выделения литотипов является выделение основных характеристик карбонатных литотипов, для каждой из которых строилась своя машина опорных векторов, а также применялся вероятностный подход, позволяющий оценить степень неопределенности прогноза.

Данная методика была опробована на реальном карбонатном месторождении Западной Сибири. Полученные результаты подтверждают гипотезу делимости литотипов в пространстве параметров и выделение литотипов для скважин без керновых исследований с ошибкой менее 10%. Полученная модель наиболее вероятного распределения литотипов по скважинам позволит более качественно строить ПДГТМ и получать более точный прогноз дебитов углеводородов. Оценка неопределенности полученных результатов необходима для расчета неопределенности прогнозов дебитов углеводородов.

ЛИТЕРАТУРА

1. *Chashkov A., Kiselev V.* Use of the Cluster Analysis and Artificial Neural Network Technology for Log Data Interpretation // *Engineering and Technologies*. 2011. 4(4).
2. *Ma Y.Z.* Lithofacies clustering using PCA and neural network: Application to wireline logs // *Mathematical Geosciences*. 2011. 43(4): 401–419.
3. *Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989.
4. *Abdi H., Williams L.J.* Principal component analysis // *Wiley Inter disciplinary Reviews: Computational Statistics*. 2010. 2: 433–459.
5. *Бабенко К.И.* Прикладная статистика: классификация и снижение размерности. Москва-Ижевск. 2002.
6. *Scholz M., Vigario R.* Nonlinear PCA: a New Hierarchical Approach // *Proceedings ESANN*. 2002. P. 433–459.
7. *Scholz M., Kaplan F., Guy C., Kopka J.* Non-linear PCA: a missing data approach // *Bioinformatics*. 2005. Vol. 20, № 21. P. 3887–3895.

8. *Garcia-Lopez F.C., Garcia-Torres M., Melian B., Moreno-Perez J.A., Moreno-Vega J.M.* Solving feature subset selection problem by a Parallel Scatter Search // *European Journal of Operational Research*. 2006. 169(2). P. 477-489.
9. *Oommen T., Misra D., Twarakavi N.K.C., Prakash A., Sahoo B., Bandopadhyay S.* An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing // *Mathematical Geosciences*. 2008. 40(4).
10. *Haykin S.* *Neural Networks: A Comprehensive Foundation*, second edition // Prentice-Hall, 1999.
11. *Platt J.C.* Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods // Cambridge, MIT Press. 1999. P. 61–74.
12. *Metropolis N., Ulam S.* The Monte Carlo Method // *J. Amer. statistical assoc.* 1949. Vol. 44, № 247, P. 335–341.