

Использование метода ближайших соседей при восстановлении обстановки осадконакопления

В. В. Белозеров¹, А. С. Бочков¹, О. С. Урмаев¹, О. М. Фукс^{1,2}
Phuks.OM@gazprom-neft.ru

¹ООО «Газпромнефть НТЦ», Санкт-Петербург, Россия, ²МФТИ, Москва, Россия

Целью данной работы является построение метода изучения геологической структуры нефтяных месторождений и создание модуля для автоматизации распознавания литолого-фациальной обстановки на основе промысловых данных каротажных диаграмм. Методами машинного обучения решается задача классификации по типам фациальных обстановок каротажных данных с месторождения с использованием спектрального представления геофизических полей. Данная методика была успешно применена к интерпретации фаций на реальном месторождении.

Ключевые слова: геостатистика; обстановка осадконакопления; фациальный анализ; спектральный метод; машинное обучение

Application of nearest neighbour method for sedimentation environment study

V. V. Belozerov¹, A. S. Bochkov¹, O. S. Ushmaev¹, and O. M. Fuks^{1,2}

¹LLC "Gazpromneft NTC", St. Petersburg, Russia, ²MIPT, Moscow, Russia

Background: Distribution properties of sedimentary rocks determine considerably the geometry and size of the reservoir and, consequently, the volume of hydrocarbon reserves. Therefore knowledge about general patterns of sedimentation rock formation is of crucial practical importance. This work suggests the method to study the geological structure of oil field by automated recognition of lithofacial environment on the basis of geophysical field data.

Methods: In the work, the spectral method is used in geophysical field representation, which is well-known for its effective application to the simulation of low-permeability and high-splitted reservoirs in the nonsteady and anisotropic conditions. The input data are geophysical data interpreted by the filed geologist, which then form the training set for the machine learning algorithm. To reduce the dimensionality of the data, only their significant features (Fourier coefficients) are retained in the learning step of the algorithm. Further, the data are classified into the different facial regions using the machine learning technique.

Results: The method was tested on the real field and with the electrometric well data as the input, it allowed to classify the wells according to the lithofacial sedimentation environment.

Concluding Remarks: In the article, the method of facial environment reconstruction is described and its applicability to the real field is shown.

Keywords: geostatistics; sedimentation environment; facial analysis; spectral method; machine learning

Введение

Как известно, аккумуляция осадков, в которых возможна генерация углеводородов, происходит при определенных физико-географических условиях. Особенности распространения осадочных пород в пространстве и во времени значительно определяют фильтрационно-емкостные свойства залежи (ФЕС), от которых в свою очередь зависят продуктивность и, в конечном счете, накопленная добыча скважин. Поэтому знание общих и частных закономерностей образования осадочных толщ имеет неоспоримое практическое значение. При создании концепции разработки месторождения основным источником информации о залежи являются данные геофизических исследований с разведочных скважин. При прогнозировании распространения свойств геофизических параметров существенным может оказаться знание характера осадконакопления, которое привело к образованию той или иной геологической обстановки. Интерпретация фаций¹ является важным этапом в определении условий седиментации обломочных осадков — сформировалось ли отложение на месте речной дельты, либо в результате миграции русла меандрирующей реки, глубоководное или мелководное это отложение и т. д. Таким образом, комплексная диагностика условий формирования осадков дает ценную информацию, необходимую для построения согласованной геологической модели месторождения и оптимальной эксплуатации залежи.

Построение палеогеографических карт, отражающих наиболее достоверные обстановки осадконакопления на стадии завершения формирования продуктивного пласта-коллектора, представляет собой чрезвычайно науко- и трудоемкую задачу. Для большинства терригенных горных пород широко распространен метод анализа каротажных диаграмм по отдельным скважинам и определения фаций на основе их электрометрических моделей [1]. Но данный вид работ требует не только глубокого понимания основ процесса осадконакопления, но и наличие качественной информации о литологическом составе изучаемых горных пород. К сожалению, наличие качественного описания кернового материала не всегда представляется возможным в виду проведения подобных работ различными компаниями в различное время. К тому же большинство материалов представлены в бумажном виде, что затрудняет и замедляет работу по обработке исходной информации. С целью ускорения и оптимизации работы геолога и предлагается данная методика для распознавания литолого-фациальной обстановки месторождения.

Результативность и прогнозная способность производственных решений при разработке нефтяных месторождений во многом зависит от качества исходной геологической модели. В основе почти всех современных пакетов геологического моделирования лежат классические методы геостатистики [2, 3], главным из которых является вариограммный метод. При этом предположения классической геостатистики — это стационарность и изотропность случайного поля геофизического параметра (а также их ослабления: тренд и стационарные остаток и геометрическая анизотропия) [4]. Однако ни стационарность, ни изотропность в своих строгих формах почти не встречаются при работе с реальными данными. В настоящее время будущее нефтяной отрасли состоит в освоении низкопроницаемых и высокорасчлененных коллекторов, применение к которым вариограммного метода дает заведомо ложный результат [5]. Поэтому для обеспечения высоких технологических

¹Фация — это тело горной породы со специфическими особенностями. При описании осадочной породы фация может быть выделена по цвету, характеру слоистости, составу, структуре, ископаемым остаткам, осадочным текстурам. Отбор признаков для определения фации и вес, присваиваемый каждому из них, зависят от субъективной оценки геолога.

экономических показателей при моделировании таких коллекторов в условиях нестационарности, зональной изменчивости и анизотропности геофизических полей требуются другие методы, лишенные ограничений классической геостатистики. Одним из таких методов является спектральный подход в представлении геофизического поля [5–7]. Как правило, каротажные данные с месторождения представляют собой сильно и случайным образом флуктуирующие функции. Пример каротажных данных приведен на рис. 1.

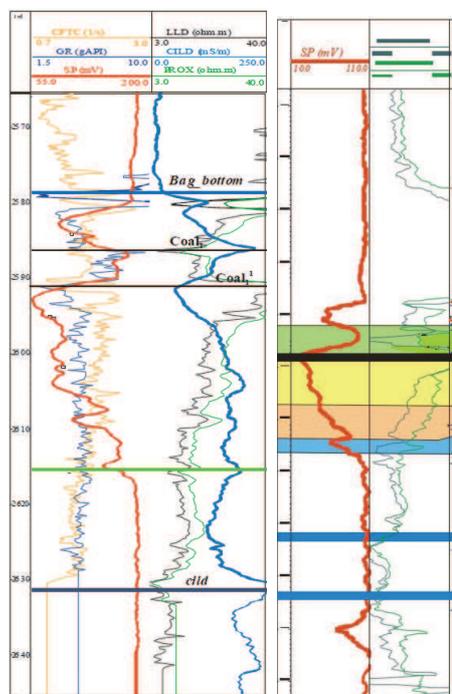


Рис. 1. Пример каротажных данных: по оси y откладывается обычно глубина, по оси x — значения каротажа

Далее статья организована следующим образом. В первом разделе описаны постановка задачи и цель работы. Затем речь идет об используемых методах решения — спектральном методе представления геофизических полей, выборе информативных признаков для описания каротажных данных и об алгоритме машинного обучения. Последний раздел содержит пример опробования методики для классификации фаций на участке реального месторождения и анализ ошибки предложенного алгоритма.

Постановка задачи

Как было упомянуто ранее, данные геофизических исследований являются основным источником информации о продуктивном пласте на стадии построения геологической модели и создании концепции разработки месторождения. Различные формы каротажных данных свидетельствуют о различных условиях осадконакопления и, соответственно, условиях формирования коллектора. Целью данной работы является разработка метода распознавания литолого-фациальной обстановки по каротажным данным.

В терминах машинного обучения задача распознавания фациальных обстановок относится к задаче классификации — необходимо разделить множество объектов X (множество каротажных данных с месторождения) на M непересекающихся классов из Y (различные типы обстановок осадконакопления). Таким образом, определены понятия пространства

объектов и пространства классов. Как известно, в задачах машинного обучения выделяют два этапа — этап обучения и этап применения. В данном случае обучающая выборка представляет собой набор проинтерпретированных геологом каротажных кривых, где для каждого элемента известно, к какой фациальной обстановке он относится, причем в ней должен присутствовать хотя бы один элемент из каждого класса Y . Неотъемлемым подготовительным этапом для работы алгоритма классификации является также отбор признаков объектов, о котором речь пойдет далее.

Метод решения

Спектральный метод. В задаче распознавания фациальной обстановки спектральный метод представления геофизических полей заключается в разложении каротажных данных (далее для удобства будем называть их также функции каротажа) в ряд Фурье. Пусть $f(h)$ — нормированная функция каротажа в некоторой скважине в отбитом маркерами пласте, переведенная линейным преобразованием на интервал $[-\pi; \pi]$, здесь h — вертикальная координата (глубина). Выявление характерной формы каротажных данных осуществим посредством разложения функции $f(h)$ по базису периодических функций, ортогональных на промежутке $[-\pi; \pi]$:

$$f(h) = S_n(h) + \Delta(h), \quad (1)$$

где $\Delta(h)$ — некоторый остаток;

$$S_n(h) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kh + b_k \sin kh). \quad (2)$$

Коэффициенты разложения a_k и b_k могут быть найдены следующим образом:

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(h) dh; \quad (3)$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(h) \cos kh dh, k \in [1; n]; \quad (4)$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(h) \sin kh dh, k \in [1; n]. \quad (5)$$

Итак,

$$f(h) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kh + b_k \sin kh) + \Delta(h). \quad (6)$$

При этом порядок разложения n определяется заданной точностью, с которой необходимо приблизить функцию каротажа $f(h)$ суммой гармонических функций.

Выбор информативных признаков. Далее для решения задачи классификации в рамках методов машинного обучения необходимо задать признаковое описание объектов. Кроме того, оно должно быть выбрано таким образом, чтобы было возможным эффективно дифференцировать объекты из различных классов. В ходе решения задачи

классификации по типам фациальных обстановок в качестве признаков объектов используется совокупность коэффициентов $\{a_n, b_n\}$ разложения функции каротажа в ряд Фурье. Выбор информативных признаков осуществляется посредством определения порядка разложения функции в ряд необходимого для корректного описания каротажных данных.

Из анализа амплитудных спектров различных каротажных данных было получено, что средняя амплитуда гармоник затухает с ростом частоты (рис. 2), т. е. убывает средняя энергия гармоник. Исходя из этого, для восстановления сигнала и в том числе для эффективного снижения размерности задачи предлагается следующая методика: оставить лишь n первых гармоник, а n следует выбрать таким образом, чтобы остаток, состоящий из высокочастотных гармоник, удовлетворял бы требованию, что средняя (по всем скважинам) относительная ошибка восстановления каротажа ε меньше изначально заданной постоянной величины ε_{max} (это есть та точность, которая необходима для корректного восстановления данных). При этом сама относительная ошибка восстановления каротажа ε для одной скважины при некотором порядке разложения функции в ряд n может быть найдена следующим образом:

$$\varepsilon = \frac{\int_{-\pi}^{\pi} \Delta^2 dh}{\int_{-\pi}^{\pi} f(h)^2 dh} = \frac{\int_{-\pi}^{\pi} (f(h) - S_n(h))^2 dh}{\int_{-\pi}^{\pi} f(h)^2 dh}. \quad (7)$$

Поэтому необходимое для спектрального разложения (6) количество коэффициентов определяется так: сначала для каждого примера данных из обучающей выборки считается относительная ошибка восстановления каротажа ε при некотором первоначальном порядке разложения n . При этом, следует заметить, что основную роль играет наличие каротажных данных со значительными неоднородностями — для их корректного восстановления потребуется больше коэффициентов. Далее берется усредненное по всем элементам обучающей выборки значение этой ошибки, и если оно превышает максимально допустимое значение ε_{max} , то количество коэффициентов увеличивается и рассчитывается относительная ошибка при новом числе коэффициентов. Так происходит до тех пор, пока средняя относительная ошибка превышает ε_{max} , т. е. пока не достигнута необходимая точность восстановления каротажных данных.

Алгоритм машинного обучения. Следующим этапом в восстановлении обстановки осадконакопления является разбиение множества каротажных данных на классы различных обстановок осадконакопления (или по-другому — фациальных обстановок). Для классификации применяется алгоритм взвешенных ближайших соседей [8], который основан на гипотезе о том, что схожие по признакам объекты принадлежат, как правило, одному классу. Для простоты задачи естественным образом вводится евклидова метрика в пространстве объектов: для двух объектов u и v , которые характеризуются совокупностью признаков $u = (u_1, \dots, u_n)$ и $v = (v_1, \dots, v_n)$ (в нашем случае это коэффициенты разложения в ряд Фурье, и их количество при этом считается заданным, оно было определено ранее), расстояние $\rho(u, v)$ между двумя этими объектами в признаковом пространстве будет

$$\rho(u, v) = \sqrt{\sum_{j=1}^n (u_j - v_j)^2}. \quad (8)$$

Чем меньше расстояние $\rho(u, v)$, тем более схожи объекты u и v .

Алгоритм взвешенных ближайших соседей заключается в том, что произвольному объекту $u \in X$ из тестовой выборки будет поставлен в соответствие класс $y \in Y$, который име-

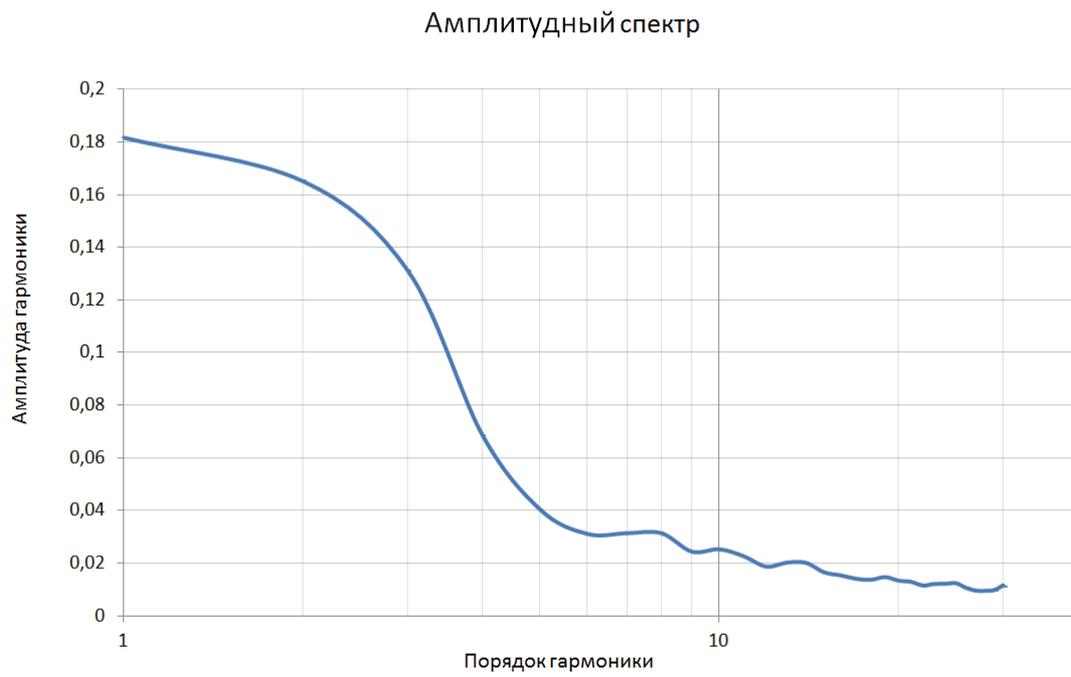


Рис. 2. Амплитудный спектр для примера каротажных данных

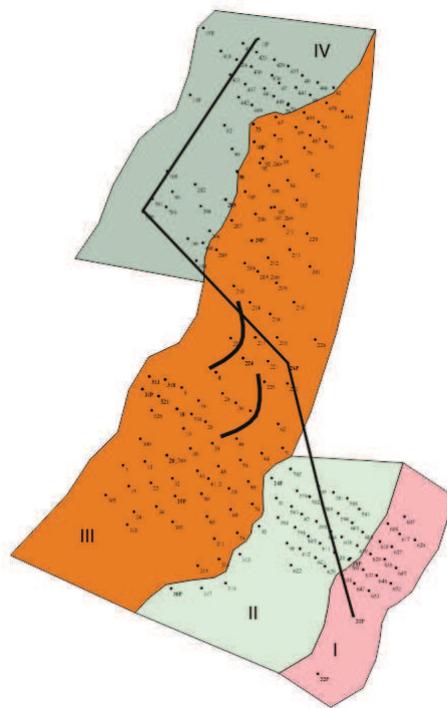


Рис. 3. Расположение скважин месторождения: различные цвета соответствуют различным выделенным фациальным зонам

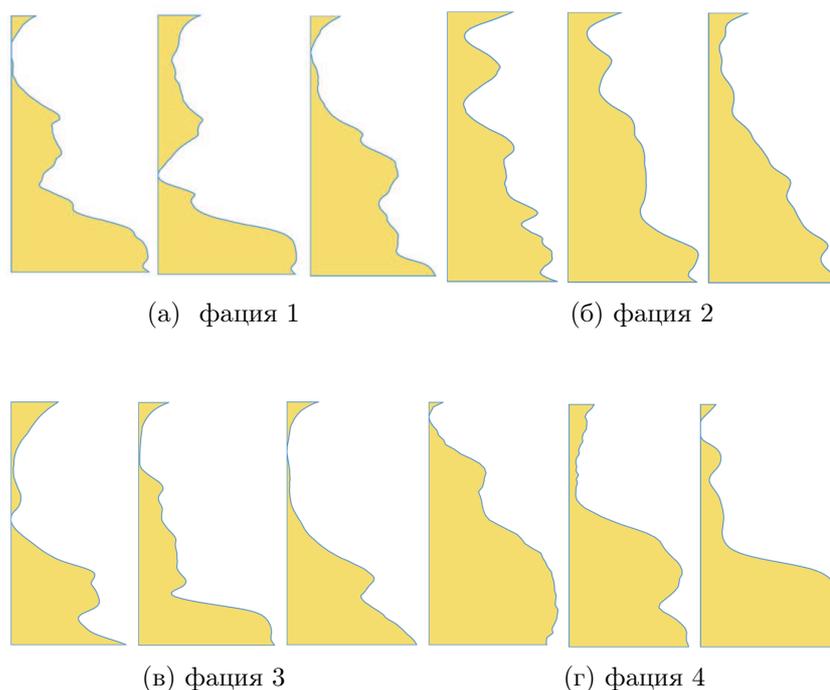


Рис. 4. Примеры данных в обучающей выборке (проинтерпретированные фации)

ет наибольший вес среди его k ближайших соседей из обучающей выборки $X_l = (x_i, y_i)_{i=1}^l$. При этом соседство определяется на основе введенной метрики ρ (8) и каждому соседу x_i объекта u присваивается соответствующий вес $w(i, u)$. Имея в распоряжении метрику пространства ρ , введем веса соседей обратно пропорционально квадрату расстояния до объекта u :

$$w(i, u) = \frac{1}{\rho^2(u, x_i)}. \quad (9)$$

Оптимальное значение параметра k , числа соседей, обычно определяется по критерию скользящего контроля с исключением объектов по одному. В ходе этой процедуры для каждого объекта $x_i \in X_l$ проверяется, правильно ли он классифицируется по своим k ближайшим соседям. Оптимальным выбирается то значение числа соседей, при котором количество ошибочных классификаций минимально. Таким образом, согласно данному алгоритму для каждого объекта каротажных данных будет определен его класс — тип обстановки осадконакопления.

Эксперимент

Описанная выше методика была применена к восстановлению обстановки осадконакопления на месторождении, которое включало 183 скважины (рис. 3). На рис. 3 различными цветами представлено истинное разбиение по фациальным зонам, которое будет использоваться в конце для анализа ошибки предложенного алгоритма. Входными данными для задачи были каротажные диаграммы, отметки кровли и подошвы пласта по каждой скважине, координаты расположения скважин на плоскости и необходимо было классифицировать все скважины по 4 различным фациальным обстановкам. Для применения алгоритма ближайших соседей была составлена обучающая выборка (рис. 4) с количеством скважин $n_L = 57$, что составляет примерно 30% от их полного числа. На рис. 4

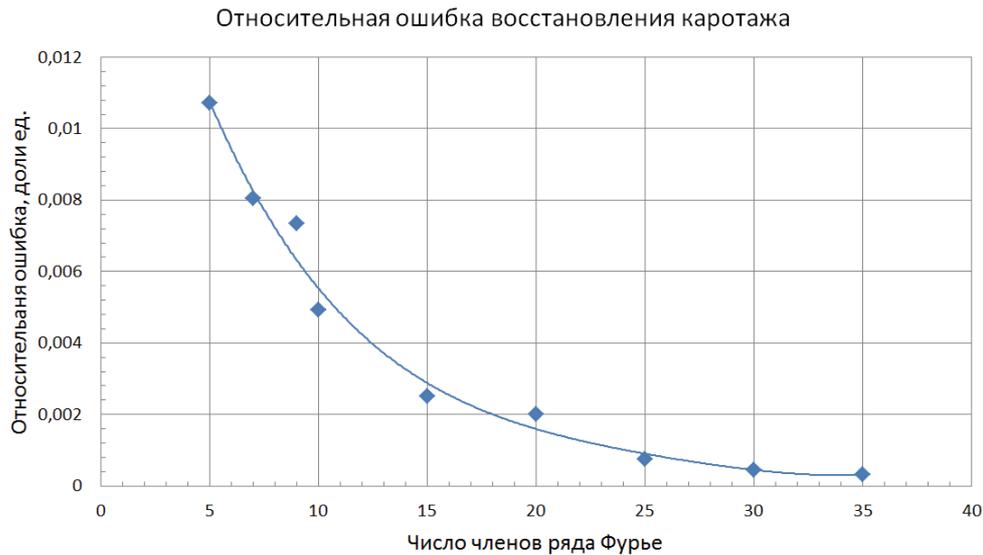


Рис. 5. Пример зависимости относительной ошибки восстановления каротажных данных ε от количества членов разложения для одной скважины

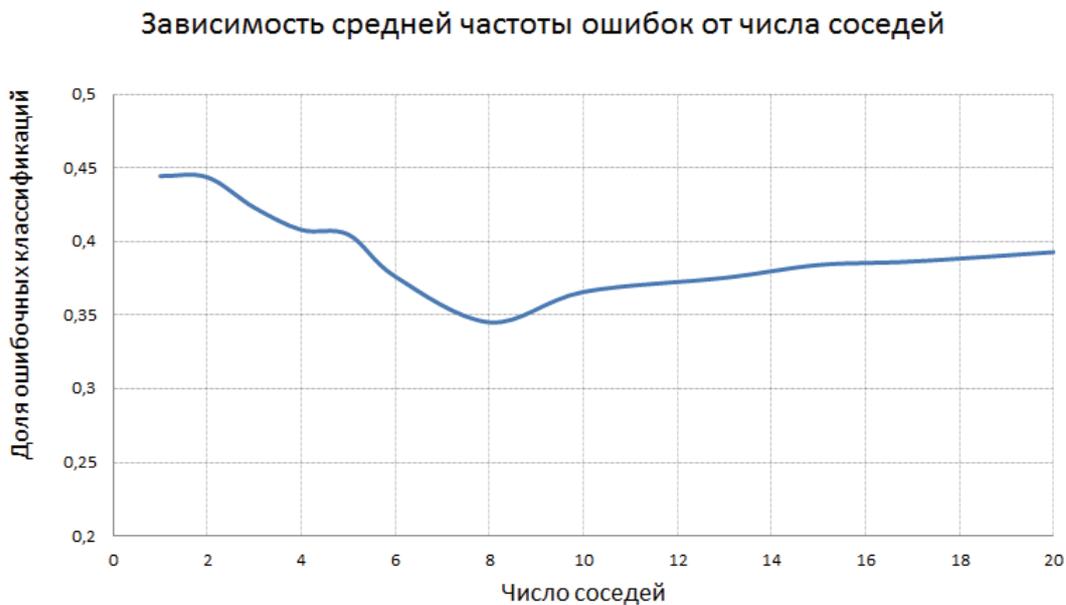


Рис. 6. Зависимость средней частоты ошибок на контрольной выборке от числа соседей, используемых в алгоритме классификации

для каждой скважины представлены следующие данные: по вертикальной оси отложена глубина, по горизонтальной — соответствующее значение каротажа на этой глубине. Для проверки точности работы алгоритма использовалась контрольная выборка, состоящая из всех оставшихся скважин, не вошедших в обучающую выборку.

Вначале был проведен анализ входных данных и определен необходимый порядок разложения функций каротажа в ряд Фурье. По формуле (7) был проведен расчет относительной погрешности восстановления каротажа для каждой скважины. В результате было получено, что при порядке разложения $n = 12$ средняя по всем скважинам ошибка восстановления каротажа $\varepsilon < 0,02$ доли ед., и в дальнейших расчетах использовался именно этот порядок разложения каротажных данных в ряд Фурье. Пример зависимости относительной ошибки восстановления каротажных данных для одной скважины от количества членов ряда Фурье показан на рис. 5.

Далее для классификации скважин был применен алгоритм взвешенных ближайших соседей, описанный подробно в предыдущем разделе. При этом оптимальное значение числа соседей было определено по критерию скользящего контроля следующим образом: для различных значений числа соседей вычислялась частота ошибочных классификаций алгоритма на контрольной выборке, которая затем была усреднена по различным наборам обучающих выборок одинаковой длины. Результаты приведены на рис. 6, из которого следует, что оптимальное число соседей для работы алгоритма на этих данных равно $n_{opt} = 8$. При этом в результате работы алгоритма была получена следующая классификация скважин, представленная на рис. 7, где различные цвета соответствуют различным выделенным обстановкам осадконакопления. Полученное таким образом площадное распределение выделенных фаций в целом, как видно из сравнения рис. 3 и рис. 7, согласуется с морфологией песчаных тел на выбранном участке. Средняя точность алгоритма на контрольных данных при этом составила около 65%.

Заключение

В статье приведена методика распознавания фациальной обстановки, которая была успешно применена к данным реального месторождения. На основе обучающей выборки в результате работы алгоритма машинного обучения были выделены фациальные зоны, при этом процент правильных классификаций на контрольной выборке составил в среднем 65%. В дальнейшем для повышения эффективности распознавания на месторождениях с более сложной геологической структурой планируется реализация более мощных алгоритмов машинного обучения.

Литература

- [1] *Муромцев В. С.* Электрометрическая геология песчаных тел — литологических ловушек нефти и газа. М: Недра, 1984. 260 с.
- [2] *Matheron G.* The theory of regionalized variables and its applications. Fontainebleau: Center of Geostatistics, 1971. 212 p.
- [3] *Matheron G.* The intrinsic random functions and their applications // *Adv. Appl. Probability*, 1973. Vol. 5. P. 439–468.
- [4] *Дюбрюль О.* Использование геостатистики для включения в геологическую модель данных. European Association of Geoscientists and Engineers, 2002. 295 с.
- [5] *Байков В. А., Бакиров Н. К., Яковлев А. А.* Новые подходы в теории геостатистического моделирования // *Вестник УГАТУ*, 2010. Т. 37, № 2. С. 209–215.
- [6] *Бочков А. С., Мухаммадеев Д. С.* Новые подходы геостохастического моделирования месторождений // *Доклад ООО «РН-УфаНИПИнефть»*, 2011.

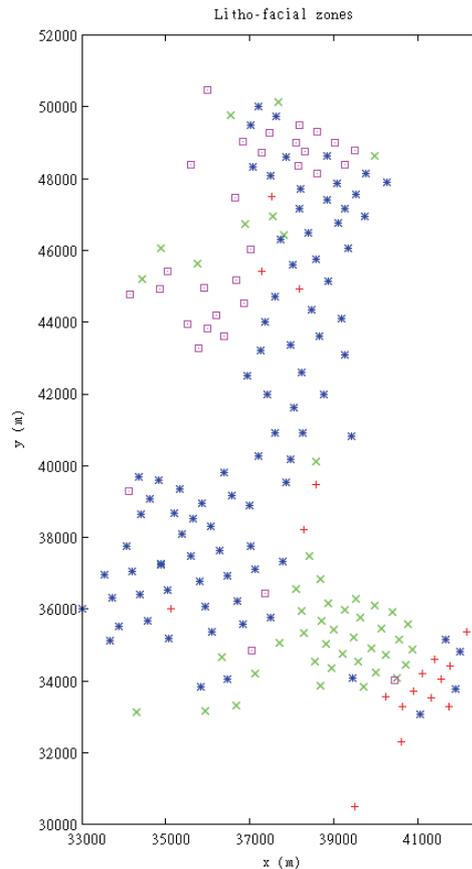


Рис. 7. Площадное распределение фаций

- [7] Байков В. А., Бочков А. С., Яковлев А. А. Учет неоднородности при геолого-гидродинамическом моделировании Приобского месторождения. Нефтяное хозяйство, 2011.
- [8] Cover T., Hart P. Nearest neighbor pattern classification // *IEEE Trans. Inform. Theory*, 1967. Vol. 13, No. 1. P. 21–27.

References

- [1] Muromtsev, V. S. 1984. Electrometric geology of sand bodies — lithological traps of oil and gas. Moscow: Nedra. 260 p. (in Russ.)
- [2] Matheron, G. 1971. The theory of regionalized variables and its applications. Fontainebelau: Center of Geostatistics. 212 p.
- [3] Matheron G. 1973. The intrinsic random functions and their applications. *Adv. Appl. Probability* 5:439–468.
- [4] Dubrule, O. 2002. Application of geostatistics for geological data model. European Association of Geoscientists and Engineers. 295 p. (in Russ.)
- [5] Bajkov, V. A., Bakirov N. K., Yakovlev A. A. 2010. New methods in the theory of geostatistical modelling. *Vestnik USATU* 37(2):209–215. (in Russ.)
- [6] Bochkov, A. S., Mukhamadeev D. S. 2011. New approaches to geostochastic modelling of oil fields. *Report RN-UfaNIPIneft LLC*. (in Russ.)

-
- [7] *Bajkov V. A., Bochkov A. S., Yakovlev A. A.* 2011. Accounting of nonhomogeneity in Priobskoye field geological modelling and simulation. Oil industry. (in Russ.)
- [8] *Cover T., Hart P.* 1967. Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* 13(1):21–27.