

Т.Ш. КАЛЬМЕНОВ, Ф.А. МУРЗИН**,
Н.В. ПОПЛЕВИНА**, Д.Ф. СЕМИЧ***

АНАЛИЗ ДАННЫХ РАДИОАКТИВНОГО КАРОТАЖА НА ОСНОВЕ МЕТОДА И.Б. МУЧНИКА

*Институт математики, информатики и механики КН МОН РК

**Институт систем информатики им. А.П. Ершова СО РАН, г. Новосибирск

1. Введение

С развитием нефтедобывающей отрасли идёт развитие методов исследования скважин и интерпретации каротажных данных. В 2003 году Новосибирским ОКБ совместно с Киевским ОКБ ОАО «Западно-Сибирская Корпорация Тюменьпромгеофизика» (в дальнейшем – ЗСК ТПГ) был разработан прибор импульсного нейтронного гамма-каротажа ИНГК-С-95. Метод исследования данным прибором – С/О каротаж – используется для оценки нефтенасыщенности в обсаженных скважинах для определения литологии и пористости.

По заказу ЗСК ТПГ Новосибирским ОКБ совместно с ИСИ СО РАН была начата разработка метода интерпретации первичных каротажных данных, полученных с прибора. В результате была разработана библиотека алгоритмов обработки исходных спектров, на основе которой создана программа «Анализатор спектров», которая сейчас активно используется. Результатом работы «Анализатора спектров» является около 2-х десятков аналитических параметров (химических интерпретационных индексов и др.). Наиболее информативными считаются три из них: отношение С/О (по спектру ГИНР), отношение Ca/Si (по спектру ГИРЗ или по спектру ГИНР) и пористость. ГИНР – гамма-излучение наведенной радиоактивности. ГИРЗ – гамма-излучение радиоактивного захвата.

Были предложены несколько алгоритмов вычисления нефтенасыщенности на основе этих трёх параметров, а также создано соответствующее программное обеспечение [1,2]. Алгоритмы основаны на усовершенствованных методах «Кросс-плот» и «Дельта С/О», которые легли в основу программного обеспечения OilTemper. Необходимо отметить, что методы «Кросс-плот» и «Дельта С/О» используются в работе ведущих сервисных геофизических компаний мира, в том числе Schlumberger и Halliburton [3,4].

Однако существуют скважины, имеющие очень плохую структуру. Для них корректно определить коэффициент нефтенасыщенности не удается. Так как обычно полезные ископаемые залегают пластами, то для решения поставленной задачи был применен также кластерный анализ [5]. В этом случае ставится более частная задача – просто выделить нефтенасыщенные пласти, определение конкретных значений коэффициента нефтенасыщенности уже не ставится. Анализировались типовые данные С/О, Ca/Si и пористость.

Однако в результате проведения радиоактивного каротажа получается большое количество аналитических параметров, которые в настоящее время слабо вовлечены в процесс интерпретации данных. Например, кроме упомянутых выше С, О, Ca, Si, определяются также интерпретационные химические индексы (монотонно зависящие от концентраций), соответствующие множеству других элементов: H, B, K, Cl, Fe, U, Th и др. Для того чтобы кластеризовать эти данные, авторы предложили использовать метод И.Б. Мучника [6,7], названный им “Лингвистический анализ экспериментальных кривых”. Этот метод позволяет анализировать очень сложные кривые.

2. Анализ данных радиоактивного каротажа

Лингвистический подход заключается в том, что для анализируемого типа образов создаются словарь признаков и грамматика, с помощью которой из слов словаря строятся конструкции – описания образов. Другими словами, создается специальный язык описания образов. После этого разрабатываются алгоритмы, которые на базе созданного языка анализируют конкретные объекты – реализация рассматриваемых образов.

Экспериментальные кривые являются типичной формой представления результатов технических и естественнонаучных экспериментов. Сейсмограммы и каротажные кривые в геофизике, электро- и фонокардиограммы, электроэнцефалограммы, электромиограммы в медицине и физиологии, записи шумов и вибраций машин в технической диагностике, инфракрасные спектры коле-

баний молекул в молекулярной спектроскопии, масс-спектры, осцилограммы речи – вот небольшой перечень видов экспериментальных кривых.

Подавляющее большинство экспериментальных кривых обладает внутренней структурой. Их анализ неизбежно связан с необходимостью выделения сравнительно коротких фрагментов, несущих отдельные единицы информации об исследуемом объекте, изучения свойств и классификации этих фрагментов и закономерностей их чередования на кривой.

При анализе данных радиоактивного каротажа, с помощью лингвистического подхода нам интересно формирование алфавита, каждый символ которого соответствует определенному элементарному событию. Символы получившегося алфавита отображаются в цветовое пространство с целью визуального выделения слоёв по глубине. Перед применением лингвистического анализа возможна подготовка кривой для избавления от шумов с помощью преобразования Хаара.

В итоге для анализа данных используется несколько методов.

1. Лингвистический анализ (включает в себя сегментацию кривой, построение алфавита и построение грамматики).

2. Преобразование Хаара.

3. Отображение в цветовое пространство.

2.1. Сегментация

Сегментация – это выделение участков, несущих единицы информации об анализируемом процессе.

Кривая разбивается на участки, которые могут перекрывать друг друга. Для каждого участка рассчитывается функция сложности. Сложными считаются локально экстремальные участки (содержащие максимумы или минимумы в зависимости от выбранной функции).

Пусть $f^j = (f_1^j, \dots, f_l^j)$ – вектор отсчетов экспериментальной кривой на участке ω_j , $g^j = (g_1^j, \dots, g_k^j)$ – вектор признаков, представляющих собой описание поведения кривой на ω_j . (k не обязательно равно l).

Тогда в качестве функции сложности $\phi(\omega_j)$ выберем функцию сходства вектора g^j с соседними векторами, т. е. среднее значение скалярного произведения вектора g^j с его соседями:

$$\phi(\omega_j) = \frac{1}{2}[(g^j, g^{j-1}) + (g^j, g^{j+1})] = (x^j, g^j), \quad x = \frac{1}{2}(f^{j-1} + f^{j+1}). \quad (1)$$

В качестве признаков, например, выбираем величины:

$$g_s^j = \frac{f_s^j - \bar{f}^j}{\|f^j\|}, \quad s = 1, \dots, l, \quad (2)$$

где через \bar{f}^j и $\|f^j\|$ обозначены числа

$$\bar{f}^j = \frac{1}{l} \sum_{s=1}^l f_s^j, \quad \|f^j\| = \sqrt{\sum_{s=1}^l (f_s^j - \bar{f}^j)^2}. \quad (3)$$

2.2. Этап формирования алфавита

Целью автоматической классификации выделенных участков является построение алфавита, каждый символ которого соответствует определенному элементарному событию в составе изучаемого процесса. Выделенные участки рассматриваются как следы таких событий, причем событиям одного и того же типа должны соответствовать близкие участки в смысле конструируемой меры расстояния.

Для формирования алфавита И. Б. Мучником предлагается использовать автоматическую классификацию с некоторым решающим правилом, которое позволяет отнести участок к одному из классов. Необходимо сформировать такое пространство признаков, с помощью которого можно анализировать участки разной длины.

На множестве участков кривой вводится мера $r = (x^i, x^j)$. На основе обработки небольшой части участков, выделяются набор опорных точек. Тогда форму любого участка x^i можно описать

вектором признаков $y^i = (y_1^i, \dots, y_k^i)$, в качестве которого принимается вектор его расстояний до опорных точек $y_p^i = r(x^i, z_p)$.

Набор опорных точек $\{z^1, \dots, z^k\}$ позволяет преобразовать метрическое пространство X участков разной длины в обычное k -мерное евклидово пространство Y , где можно использовать известные алгоритмы автоматической классификации.

2.3. Мера различия на множестве выделенных участков

Пусть даны

$$x^i = (x_1^i, \dots, x_{l_i}^i), \quad x^j = (x_1^j, \dots, x_{l_j}^j) \quad (4)$$

– два выделенных участка, а l_i и l_j – их длины. Будем считать векторы x^i и x^j центрированными, так что

$$\bar{x}^i = \frac{1}{l_i} \sum_{t=1}^{l_i} x_t^i = 0, \quad \bar{x}^j = \frac{1}{l_j} \sum_{t=1}^{l_j} x_t^j = 0 \quad (5)$$

В соответствии с целями этапа формирования алфавита конструируемая мера различия должна объявлять близкими участки, соответствующие однотипным элементарным событиям. Такие события вызывают появление на кривой локальных особенностей примерно одной формы.

Сравнение становится осмысленным, если участки налагаются друг на друга с некоторым сдвигом τ , компенсирующим неточность сегментации. При этом компоненте x_t^i вектора x^i соответствует компонента $x_{t+\tau}^j$ вектора x^j . Целое число τ может быть положительным, отрицательным или равным нулю.

Пусть τ – некоторый фиксированный сдвиг, $|\tau| < \max(l_i, l_j)$.

Меру расстояния $\rho(x^i, x^j, \tau)$ между векторами x^i и x^j , зависящую от сдвига x^j , естественно понимать как меру их различия на всем множестве x^i и x^j , зависящую от сдвига τ :

$$\rho(x^i, x^j, \tau) = \sum_{t \in T_i(\tau)} (x_t^i)^2 + \sum_{t \in T(\tau)} (x_t^i - x_{t+\tau}^j)^2 + \sum_{t \in T_j(\tau)} (x_t^j)^2. \quad (6)$$

В этом выражении использованы следующие обозначения для общей части x^i и x^j :

$$T(\tau) = \{t_1(\tau), t_1(\tau)+1, \dots, t_2(\tau)\}, \quad (7)$$

где $t_1(\tau) = \max(1, 1 - \tau)$, $t_2(\tau) = \min(l_i, l_j - \tau)$, причем $t_2(\tau) - t_1(\tau) + 1$ – длина $T(\tau)$, и для не перекрывающихся частей

$$\begin{aligned} T_i(\tau) &= \{1, \dots, -\tau\} \cup \{l_j - \tau + 1, \dots, l_i\} \\ T_j(\tau) &= \{1, \dots, \tau\} \cup \{l_i + \tau + 1, \dots, l_j\} \end{aligned} \quad (8)$$

где члены $\{q_1, \dots, q_2\}$ каждого из обединений $T_i(\tau)$ и $T_j(\tau)$ считаются пустыми, если $q_1 > q_2$.

Разным значениям сдвига τ соответствуют разные значения такой меры различия векторов x^i и x^j . В то же время судить о «похожести» и «непохожести» двух отрезков кривой можно только при таком их взаимном расположении, при котором они максимально совпадают по форме. Поэтому в качестве меры различия участков следует принять величину

$$r(x^i, x^j) = \min_{\tau} \rho(x^i, x^j, \tau). \quad (9)$$

В алгоритмах автоматической классификации часто удобно использовать в качестве меры различия векторов функции, удовлетворяющие аксиомам метрики [4]. Мера (6) является именно такой функцией.

Разбиение на основе сети опорных точек определяет алфавит языка описания анализируемых кривых.

Алгоритм построения сети опорных точек подробно описан статье [4].

Необходимо отметить, что алгоритм построения опорных точек, описанный в статье И.Б. Мучника, требует длительного расчёта времени, т.к. использует перебор большого количества вариантов векторов. Для реализации он требует оптимизации. На данном этапе было решено заменить его алгоритмом кластеризации с метрикой, предложенной И.Б. Мучником, который позволяет также разбить на внутренне схожие группы.

В результате применения алгоритмов сегментации и автоматической классификации выделенных фрагментов экспериментальная кривая оказывается представленной в виде упорядоченной последовательности символов конечного алфавита (цепочки). Такую последовательность можно трактовать как текст на некотором языке, в данном случае языке, специально приспособленном для описания обрабатываемых кривых. При этом каждую цепочку можно рассматривать как некоторую «фразу» (или несколько фраз) на этом языке. На основе алфавита И.Б. Мучник строит грамматику, с другой стороны, алфавит форм выделенных участков служит классификатором особых событий, возникающих в ходе развития изучаемого процесса. Поэтому он может служить самостоятельным средством анализа экспериментальных кривых.

3. Программная реализация

Лингвистический анализ был встроен в программу кластеризации «Clustering», которая была реализована ранее. Приложение позволяет анализировать данные, полученные в результате радиоактивного каротажа. Данные загружаются в геофизическом формате LAS из программы «Анализатор спектров» или других аналогичных программ. В качестве данных для тестирования использовались реальные данные со скважин. Результаты анализа отображаются в цветовое пространство – визуально выделяются пласти.

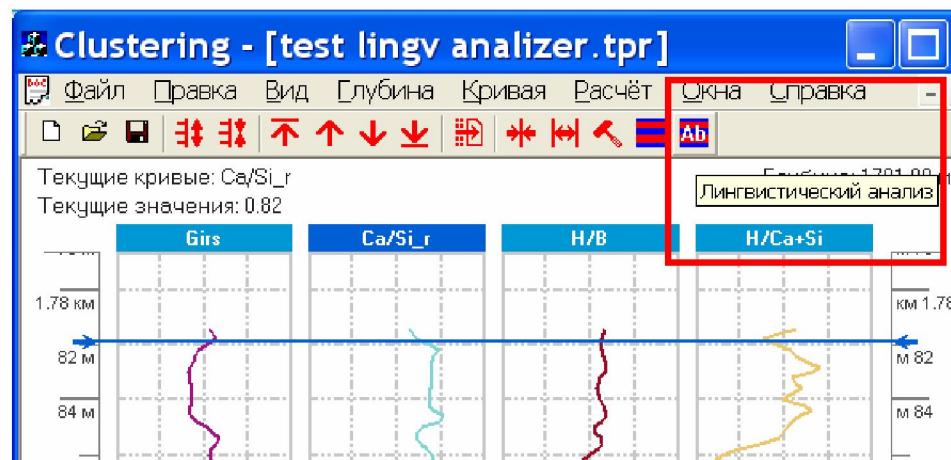


Рис. 1. Кнопка вызова диалогового окна настройки алгоритма на панели

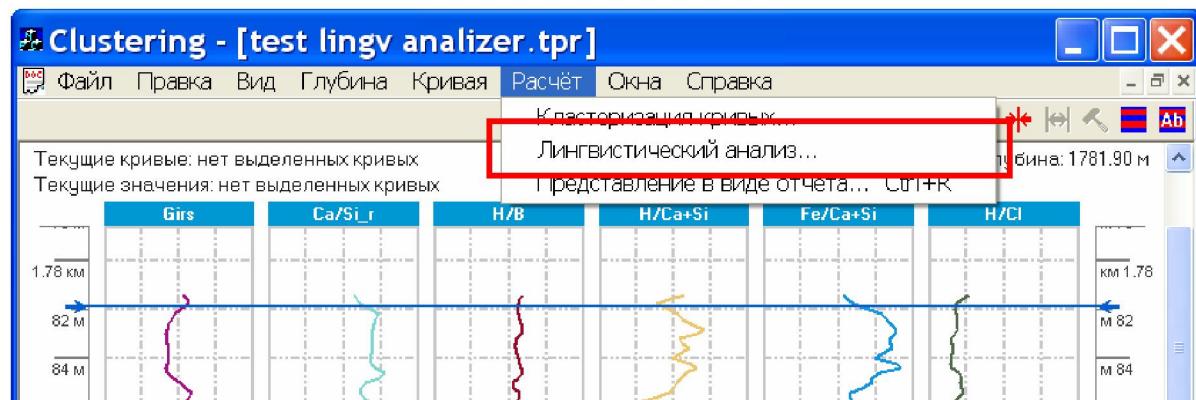


Рис. 2. Пункт меню для диалогового окна настройки алгоритма

4. Преобразование Хаара:

При анализе данных радиоактивного каротажа может оказаться полезной фильтрация высокочастотных составляющих. Для этого используется преобразование Хаара с матрицей размерностью 32×32 . Преобразование осуществляется умножением матрицы Хаара на вектор данных соответствующей длины.

В результирующем векторе зануляются коэффициенты, соответствующие гармоникам, которые считаются шумом, после чего вектор восстанавливается по скорректированному спектру. В приложении реализована возможность задавать окна на спектре для гармоник, которые хотим удалить.

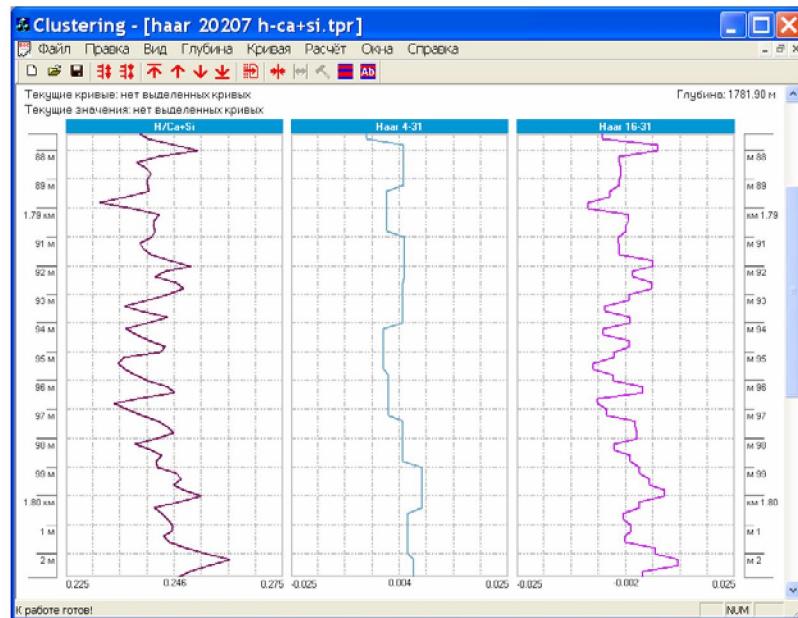


Рис. 3. Преобразование Хаара, реализованное в программе Clustering

5. Отображение в цветовое пространство

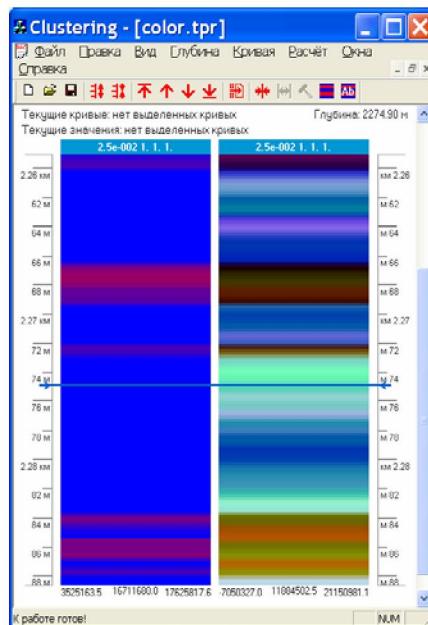


Рис. 4. Слева – планшет, построенный методом кластеризации; справа – планшет, построенный с помощью субтрактивного смешивания цветов

Для произвольного набора кривых реализована возможность создания графического планшета по глубине. Планшет с помощью цветов передаёт пропорциональное содержание химических индексов в каждой точке измерения. Каждой кривой назначается свой цвет и производится субтрактивное смешивание цветов для каждой точки глубины.

В зависимости от значения индекса в точке измерения для цвета задаётся величина альфа-канала.

Очевидна проблема, состоящая в том, что уникальных цветов только три, поэтому, чем больше кривых будет задействовано при анализе, тем больше вероятность получить при смешивании цветов не совсем корректную визуализацию.

Но для анализа небольшого количества кривых, и как вспомогательный инструмент, (если кривых больше трёх) данный способ визуализации полезен.

6. Тестирование

Было проведено тестирование программы на скважинных данных и сравнение результатов, полученных по программе, с данными, полученными «вручную» экспертами в интерпретационном центре.

Тестирование показало, что алгоритм позволяет анализировать сложные кривые, разбивая их участки на классы, которые затем можно отображать в цветовое пространство для выявления насыщенных пластов.

При тестировании сравнивались результаты, полученные методом кластеризации и методом лингвистического анализа. Тесты показали, что на некоторых скважинах качество работы алгоритмов примерно одинаковое. На некоторых скважинах качество результатов, полученных лингвистическим анализом, выше.

Например, анализировались кривые H/B, H/Ca, Fe/Ca+Si, H/Cl посредством лингвистического анализа. Затем результаты анализа сравнивались с эталонными данными, полученными из интерпретационного центра о нефтенасыщенных пластах и с результатами кластеризации. Цель такого сравнения состояла в том, чтобы понять не «сигналят» ли данные о тех или иных химических элементах о нефтенасыщенных пластах. Очевидно, что на разных месторождениях разные элементы могут «сигналить», т.е. являться индикаторными.

7. Выводы

В результате проведения радиоактивного каротажа получается большое количество аналитических параметров, которые в настоящее время слабо вовлечены в процесс интерпретации данных. Например, кроме упомянутых выше C, O, Ca, Si, определяются также интерпретационные химические индексы (монотонно зависящие от концентраций), соответствующие множеству других элементов: H, B, K, Cl, Fe, U, Th и др. Для того чтобы анализировать эти данные, было решено использовать метод И.Б. Мучника, названный им “Лингвистический анализ экспериментальных кривых”.

Было произведено встраивание соответствующей программной компоненты в реализованную ранее программу «Clusteruing». Метод был протестирован на реальных скважинных данных. Результаты в некоторых случаях (на некоторых скважинах, для некоторых химических индексов) сравнимы, а в некоторых случаях качественнее результатов кластеризации для сложных кривых. Плюсом метода является хорошее качество выделения пластов для различных данных при фиксированных параметрах алгоритма, что позволяет один раз для данного месторождения настроить параметры и в дальнейшем использовать метод, не меняя параметров алгоритма.

ЛИТЕРАТУРА

1. Винокуров А.А., Ильин И.В., Мурзин Ф.А., Семич Д.Ф. Расчет коэффициента нефтенасыщенности по данным, полученным аппаратурой ИНГК-С-95 ЗСК «Тюменьпромгеофизика» // КАРОТАЖНИК – Научно-технический вестник, (Министерство природных ресурсов РФ, Ассоциация научно-технического и делового сотрудничества по геофизическим исследованиям и работам в скважинах, Евро-Азиатское геофизическое общество), № 12-13 (125-126), 2004. – С. 41-46

2. Поплевина Н.В. Визуализация данных радиоактивного каротажа и расчет коэффициента нефтенасыщенности методом «Кросс-плот» // Конференция-конкурс «Технологии Microsoft в теории и практике программирования», Новосибирск, НГУ, фев.24-26, 2007. – С. 226-227.

3. Xu Jinwu, Zhang Zongjian. *Improved Carbon/Oxygen Log Interpretation Techniques under Variable Formation Water Salinity*. Shengli Well Logging Co., China, December 1999, 12p.

4. Джекобсон Л.А., Эттидж Р., Симпсон Дж. *Новый прибор малого диаметра с высокими характеристиками для мониторинга продуктивных пластов*. *Hulliburton Energy Services*, 1994, 14с.
5. Мандель И. Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 186с.
6. Моттель В.В., Мучник И.Б. Лингвистический анализ экспериментальных кривых // ТИИЭР, том 67, № 5, 1979. – С 12-39.
7. Muchnik I.B., "Simulation of process of forming the language for description and analysis of the form of images," *Pattern Recognition*, pp. 101-140, 1972.

REFERENCES

1. Vinokurov A.A., Il'in I.V., Murzin F.A., Semich D.F. *Nauchno-tehnicheskij vestnik*, 12-13 (125-126), **2004**, 41-46 (in Russ.).
2. Poplevina N.V. *Konferencija-konkurs Tehnologii Microsoft v teorii i praktike programmirovaniya*, Novosibirsk, NGU, 24-26 Fevral', **2007**, 226-227 (in Russ.).
3. Xu Jinwu, Zhang Zongjian. *Shengli Well Logging Co., China*, December **1999**, 12 p.
4. Dzhekobson L.A., Jetridzh R., Simpson Dzh. *Hulliburton Energy Services*, **1994**, 14 p. (in Russ.).
5. Mandel' I. D. *Finansy i statistika*, **1988**, 186 p. (in Russ.).
6. Mottl' V.V., Muchnik I.B. *TIIER*, 67, 5, **1979**, 12-39 (in Russ.).
7. Muchnik I.V., *Pattern Recognition*, **1972**, 101-140.