

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ

УДК 681.5.03

О. М. ВЕРХОТУРОВА

СИСТЕМА ВЫДЕЛЕНИЯ 3D-ОБЪЕКТОВ
НА ОСНОВЕ ДИСКРЕТНЫХ ДАННЫХ

В статье рассматривается подход к решению задачи выделения естественных неструктурированных объектов в заданной области $O \subset R^3$. Критерием их выделения в исследуемых узлах служит принцип подобия сегментов кривых, характеризующих каждый опорный и исследуемый узел. Применяются такие методы решения, как полный перебор, генетический алгоритм, метод динамического искажения времени (DTW). Все разработки велись в среде MATLAB. Подход был опробован для нахождения границ нефтенасыщенных пластов на нескольких месторождениях. *3D-объекты ; дискретные данные ; полный перебор ; генетический алгоритм ; метод динамического искажения времени*

Задача выделения объектов встречается в различных областях человеческой деятельности, в том числе в космических исследованиях, в нефтяной промышленности, в медицине и т. д. Особый интерес представляет задача выделения естественных неструктурированных объектов (например, задача выделения продуктивных пластов на скважинах месторождения), которая является весьма трудоемкой, сложной и плохо формализуемой. Обзор отечественной и зарубежной литературы, информационных интернет-источников позволяет сделать вывод, что существует несколько методов ее решения. Один из первых подходов был разработан Ш. А. Губерманом [4, 5]. Методы решения данной задачи развивались коллективом во главе с И. С. Гутманом [6]. Представляет интерес работа группы В. А. Тененева, в которой для нахождения решения использовались нейронные сети [7]. Анализ существующих методов выявил, что они либо нуждаются в трудоемкой настройке и адаптации к каждому конкретному набору входных данных, либо не учитывают ряд важных факторов, касающихся выбора исходной информации, геометрического расположения и др. Вышесказанное определяет необходимость разработки нового подхода для повышения эффективности решения задачи выделения естественных неструктурированных объектов.

1. ПОСТАНОВКА ЗАДАЧИ

Для начала введем различные обозначения.

Даны:

Рассматриваемая область – $O \subset R^3$.

Опорные узлы –

$$P_m(x_m, y_m, z_m), m = \overline{1, M}, P_m \in O.$$

Исследуемые узлы –

$$P_n(x_n, y_n, z_n), n = \overline{1, N}, P_n \in O.$$

Следует отметить, что начальное значение координаты z у каждого узла может быть свое, но поскольку с точки зрения положения на плоскости XOY это не имеет значения, то все узлы выравняются на уровень, соответствующий $z = 0$: $z_m = 0, m = \overline{1, M}$ и $z_n = 0, n = \overline{1, N}$.

Каждому опорному и исследуемому узлу соответствует несколько характеризующих его данных $g_l^m(h)$ и $g_l^n(h')$, $l = \overline{1, L}$,

h – глубина или номер замера (дискретная величина), $h \in H^m$,

h' – глубина или номер замера (дискретная величина), $h' \in H^n$.

H^m – множество глубин или множество номеров замеров для m -го опорного узла,

H^n – множество глубин или множество номеров замеров для n -го исследуемого узла.

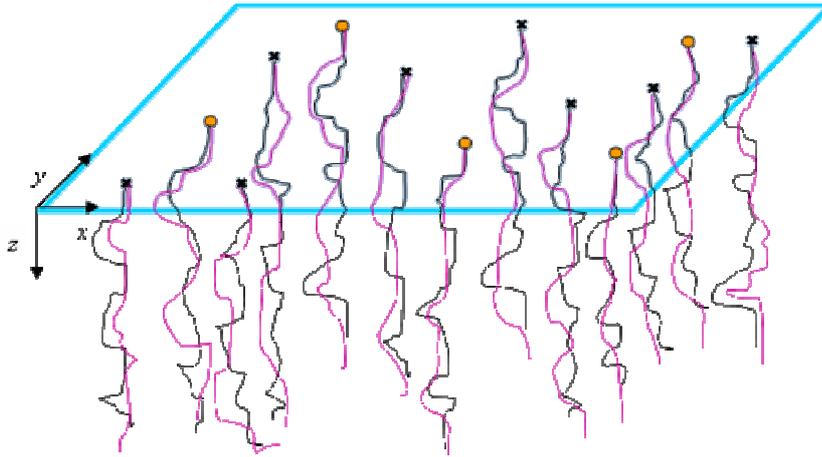


Рис. 1. Постановка задачи: опорные и исследуемые узлы (для наглядности показаны только две кривых, в общем случае L):
▭ – рассматриваемая область, O ; ● – опорный узел; ✕ – исследуемый узел

В дальнейшем $g_l^m(h)$ и $g_l^n(h')$ будут визуализироваться в виде кривых (рис. 1).

Имеется $k = \overline{1, K}$ выделенных интервалов (k объектов).

В опорных узлах границы этих интервалов известны:

- начало интервала – top_k^m – соответствующая глубина или номер замера;
- конец интервала – bot_k^m – top_k^m – соответствующая глубина или номер замера:

$$\text{top}_k^m \neq \text{bot}_k^m, \text{top}_k^m < \text{bot}_k^m.$$

Пусть G^M – множество всех кривых во всех опорных узлах, то есть все $g_l^m \in G^M$, $l = \overline{1, L}, m = \overline{1, M}$;

G^N – множество всех кривых во всех исследуемых узлах, то есть все $g_l^n \in G^N$, $l = \overline{1, L}, n = \overline{1, N}$;

Top_k^M – множество всех известных верхних границ k -го интервала в опорных узлах, то есть все $\text{top}_k^m \in \text{Top}_k^M$, $m = \overline{1, M}$;

Bot_k^M – множество всех известных нижних границ k -го интервала в опорных узлах, то есть $\text{bot}_k^m \in \text{Bot}_k^M$, $m = \overline{1, M}$

Top_k^N – множество всех неизвестных верхних границ k -го интервала в исследуемых узлах, то есть все $\text{top}_k^n \in \text{Top}_k^N$, $n = \overline{1, N}$;

Bot_k^N – множество всех неизвестных нижних границ k -го интервала в исследуемых узлах, то есть все $\text{bot}_k^n \in \text{Bot}_k^N$, $n = \overline{1, N}$.

Нужно найти границы интервалов в исследуемых узлах Top_k^N и Bot_k^N , минимизирующие целевую функцию

$$F(G^M, G^N, \text{Top}_k^M, \text{Bot}_k^M, \text{Top}_k^N, \text{Bot}_k^N) \rightarrow \min, \\ k = \overline{1, K}.$$

В качестве целевой функции может выступать, например, сумма квадратов расстояний между выделенными интервалами в опорных и исследуемых узлах (по точкам соответствующих кривых).

В вышеизложенной постановке рассматриваемая задача относится к классу задач оптимизационного геометрического моделирования.

2. СТРУКТУРА РЕШЕНИЯ

Решение можно разбить на несколько последовательных этапов:

- 1) формирование кривой для нахождения границ интервалов;
- 2) выбор способа сопоставления опорных и исследуемых узлов;
- 3) выбор критерия для сравнения интервалов (целевой функции);
- 4) нахождение границ интервалов в исследуемых узлах;
- 5) уточнение границ интервалов.

2.1. Формирование кривой для нахождения границ интервалов

У каждого узла имеется несколько характеризующих его кривых. Поэтому возникает вопрос: по каким именно кривым искать границы интересующих нас интервалов? Предлагается сравнивать интервалы не по исходным кривым, а по некоторой синтетической кривой, которую получаем в результате применения метода глав-

ных компонент (МГК) к данным всех узлов области (либо определенной части).

Основная идея МГК заключается в сокращении числа исходных переменных до наиболее существенно влияющих [1].

Главные компоненты z_j ($j = \overline{1, J}$) являются линейными комбинациями исходных переменных x_j ($j = \overline{1, J}$), и могут быть записаны в общем виде (1):

$$z_j = \alpha_j \cdot x_j, \quad j = \overline{1, J}. \quad (1)$$

При этом количество выделенных главных компонент в общем случае значительно меньше числа исходных переменных.

В предлагаемом подходе используется только первая главная компонента. Например, исходным кривым f_1, f_2, f_3, f_4 соответствует синтетическая кривая, определяемая по формуле (2).

$$f = \alpha_1 \cdot f_1 + \alpha_2 \cdot f_2 + \alpha_3 \cdot f_3 + \alpha_4 \cdot f_4, \quad (2)$$

где $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ – коэффициенты при первой главной компоненте, полученные с помощью МГК.

2.2. Выбор способа сопоставления опорных и исследуемых узлов

Дана область, в которой есть M опорных узлов и N – исследуемых. Как их соотносить друг с другом?

Можно выделить следующие варианты:

- парное сравнение узлов, когда с каждым опорным узлом поочередно сравниваются соседние исследуемые узлы, затем с ними сравниваются их соседние и т. д.
- интегральное сравнение узлов, когда решение находится сразу во всех исследуемых узлах на основе известных значений в опорных узлах.

Парное сравнение узлов хорошо тем, что в нем число узлов с известными границами интервалов постепенно увеличивается, кроме того, узлы удобно сравнивать между собой попарно. В работе был выбран именно этот вариант. Рассмотрим его более подробно.

При парном сравнении узлов одним из решающих моментов является построение путей обхода узлов исследуемой области, которое можно выполнить различными способами. В данной задаче предполагается, что наибольшим сходством обладают близлежащие узлы, а по мере удаления степень сходства постепенно уменьшается. С учетом этого были рассмотрены

следующие варианты построения путей обхода узлов:

- **путь максимальной корреляции.** Поиск начинается с опорного узла, вокруг него выбираются N самых близких по расстоянию узлов, для каждого из них находится решение (границы интервалов). Затем осуществляется переход к тому исследуемому узлу, у которого коэффициент корреляции с исходным узлом наибольший. Процесс повторяется до тех пор, пока для текущего узла наибольший коэффициент корреляции с соседними не станет меньше некоторого заданного порогового значения.

- **по профилям, составленным специалистами.** Сравнение узлов можно проводить по уже составленным специалистами профилям (профиль – последовательность узлов, первый из которых является опорным), согласно которым и осуществляется парное сравнение характеризующих каждый узел кривых. В этом случае также может задаваться пороговое значение, после достижения которого процесс прекращается.

- **по ячейкам.** В профилях расстояние от первого до последнего узла может быть значительным, при этом характерные особенности узлов (форма кривых на рассматриваемых интервалах) по мере движения по профилю постепенно меняются, и корреляция падает. Поэтому было решено применять так называемую площадную корреляцию, при которой вся область разбивается на ячейки Вороного, после чего строятся пути обхода узлов внутри каждой ячейки.

2.3. Выбор критерия для сравнения интервалов (целевой функции)

Для выделения интервалов необходимо выбрать какой-то критерий – числовой показатель, на основе которого можно делать вывод о схожести определенных участков сравниваемых кривых.

В качестве данного критерия могут выступать:

- сумма квадратов отклонений между участками сравниваемых кривых;

Является общепринятой мерой сходства между сравниваемыми кривыми. Чем меньше этот показатель, тем больше мера сходства;

- коэффициент корреляции между участками сравниваемых кривых.

При изучении корреляционной связи переменных различают две главные задачи [8]:

- 1) определение силы или степени связанности (*схожести*) переменных (аналитиче-

ское выражение связи не имеет значения), которая характеризуется величиной коэффициента корреляции;

2) определение формы (аналитической *зависимости*) корреляционной связи (линейная, нелинейная).

В рассматриваемом случае речь идет только о первой из них.

Значения коэффициента корреляции расположены в диапазоне от -1 до 1, чем ближе значение находится к 1, тем больше мера сходства между рассматриваемыми переменными.

- Косинус угла между участками сравниваемых кривых.

Это функция схожести, основанная на векторном представлении данных. Если косинус угла равен 1, векторы коллинеарны, что говорит о схожести участков сравниваемых кривых. Если косинус равен 0, векторы перпендикулярны, то есть кривые абсолютно несхожи. Если косинус угла между векторами имеет промежуточное значение, то кривые имеют некоторую сонаправленность, выражаемую косинусом угла между ними.

Проведенные исследования показали, что самым информативным и легко интерпретируемым является коэффициент корреляции. Именно его предлагается использовать для вычисле-

ния меры схожести определенных участков сравниваемых кривых. Применение выбранного критерия предполагает сравнение интервалов с одинаковым количеством данных, соответственно, участки сравниваемых кривых необходимо приводить к равному количеству точек.

2.4. Нахождение границ интервалов в исследуемых узлах

Процесс поиска происходит таким образом, что одновременно ищутся границы только одного интервала. Каждое решение состоит из верхней границы интервала (Top) и ширины интервала (Dh), в результате определяется нижняя граница интервала (Bot = Top + Dh).

Пространство поиска задается:

- дельтой Δ_1 – диапазоном возможных значений $[\text{Top} - \frac{\Delta_1}{2}; \text{Top} + \frac{\Delta_1}{2}]$, в пределах которого может лежать верхняя граница интервала;
- дельтой Δ_2 , в которой может находиться ширина интервала $[\text{Dh} - \frac{\Delta_2}{2}; \text{Dh} + \frac{\Delta_2}{2}]$.

Затем происходит поиск в получившемся пространстве возможных решений (рис. 2).

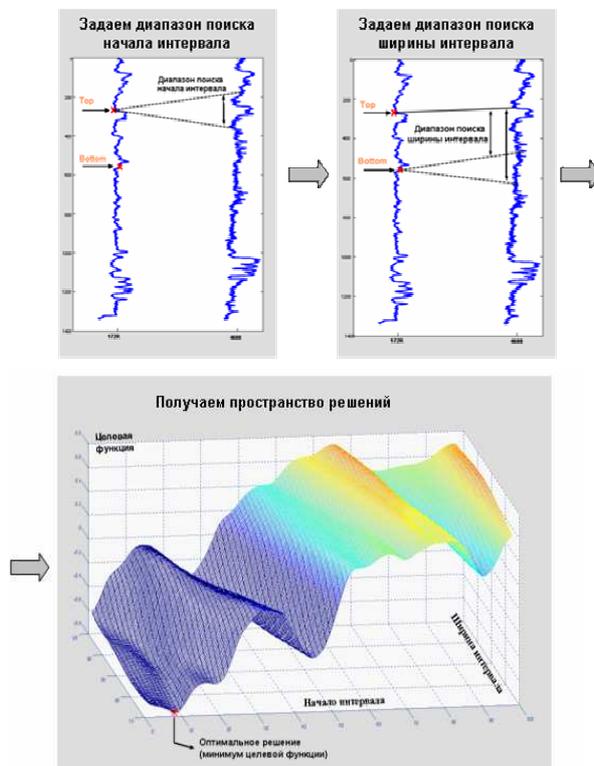


Рис. 2. Схема действий по заданию пространства решений

Для поиска решения реализовано два альтернативных метода:

- полный перебор (в случае, если пространство решений небольшое, и можно найти решение за приемлемое время);
- генетический алгоритм (ГА).

Полный перебор является точным методом, который всегда находит оптимальное решение в заданном пространстве возможных значений, но при большом диапазоне поиска требует значительных затрат времени.

Генетический алгоритм [10] является эвристическим методом, поэтому он не гарантирует нахождения оптимального решения, но более приемлем с точки зрения времени и вычислительных затрат.

Поскольку генетический алгоритм является универсальным методом, основная сложность заключается в том, что нужно правильно закодировать исходные данные и настроить алгоритм под решение данной конкретной задачи.

Рассматривались различные варианты параметров алгоритма, в том числе и различные варианты операторов кроссовера и мутации. В результате был выбран кроссовер SBX (Simulated Binary Crossover) [9].

2.5. Уточнение границ интервалов

Коэффициент корреляции позволяет находить сходные интервалы, но является не очень точным в определении границ интервалов. Поэтому для уточнения границ необходимы дополнительные критерии. Один из возможных вариантов – учет знака первой производной для определения положения границы интервала:

- если знак производной на границе положительный, то граница находится в интервале возрастания;
- если знак отрицательный, то граница находится в интервале убывания;
- если производная на границе интервала равна 0, то граница соответствует локальному экстремуму.

Таким образом, в процессе поиска определяются знаки первой производной на границах интервалов в опорном узле. Затем в исследуемом узле рассматриваются возможные положения границ только с такими же знаками первой производной. Это ограничивает пространство возможных решений.

3. Альтернативный метод решения – метод DTW

Альтернативным методом выделения интервалов является метод DTW – Dynamic Time

Warping – метод динамического искажения времени [3].

Суть метода DTW заключается в нелинейном выравнивании сопоставляемых образцов, на основе которого определяется наилучшее соответствие входного и эталонного сигналов. В рассматриваемой задаче данные опорного узла являются эталонным сигналом, а данные исследуемого узла – входным (рис. 3).

Основными областями использования метода DTW являются системы распознавания речи, системы верификации рукописной подписи, медицинская диагностика. Однако он подходит и для решения рассматриваемой задачи.

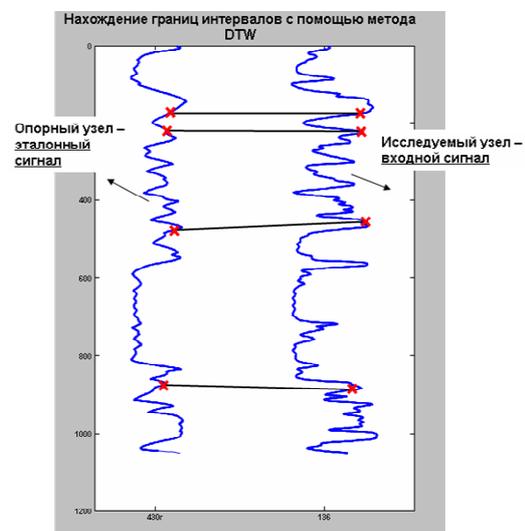


Рис. 3. Нахождение границ интервалов с помощью метода DTW

Особенно важными в этом смысле являются следующие особенности метода:

- метод хорошо подходит для сопоставления сигналов, в которых некоторая информация (некоторые сегменты) может быть пропущена;
- особенностью метода является то, что он учитывает неравномерность протекания процесса во времени.

Исходя из этого, если исследуемые интервалы имеют схожую форму кривых и хорошо выделяются по всей области, то целесообразнее применять метод, основанный на корреляции.

Если же форма кривых в исследуемых интервалах значительно изменяется, характеризуются сжатием, растяжением, отсутствием некоторых участков, то в этом случае метод, основанный на корреляции, не будет давать удовлетворительных результатов. Тогда необходимо использовать метод DTW.

4. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

Выделение геологических объектов и определение их границ составляют неперенную часть огромного количества геологических работ. В том числе весьма трудоемкой, сложной и плохо формализуемой является задача определения границ продуктивных пластов. Процесс выделения пластов в геологии носит название корреляции разрезов скважин и является одной из первых задач при построении геологической модели месторождения [2]. Следует отметить, что если другие этапы построения геологической модели уже во многом автоматизированы, то проблемы корреляции до сих пор в основном решаются методом экспертных оценок, далеких от статистической согласованности и математической строгости.

В связи с этим актуально создание компьютерной системы корреляции разрезов скважин по данным ГИС (геофизического исследования скважин), позволяющей определять границы продуктивных пластов на эксплуатационных скважинах месторождения на основе сети опорных скважин, для которых эти границы известны.

В качестве опорных скважин обычно используются разведочные скважины месторождения. При этом у каждой скважины есть различные данные ГИС, которые графически представлены каротажными кривыми.

На опорных скважинах известны границы продуктивных пластов: верхняя граница – кровля пласта, нижняя граница – подошва пласта.

Границы пластов также называются маркерами. По схожести участков каротажных кривых требуется найти эти границы на эксплуатационных скважинах месторождения.

5. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Была разработана система, которая позволяет:

- загружать необходимые данные по месторождению;
- устанавливать параметры поиска;
- сохранять результаты в удобном для геолога формате.

Полученные результаты могут быть представлены в графическом виде. Имеется возможность вывести профиль (последовательность скважин) с выделенными границами. Триангуляционно-корреляционная сеть скважин и карта толщин (рис. 4) выделяют слабые места или те скважины, которые требуют дополнительного анализа. Выводится также общий коэффициент корреляции скважин по месторождению.

В качестве результатов приведем таблицу сравнения границ пластов, полученных с помощью разработанной системы, с границами пластов экспертов (геологов) (табл. 1). Из таблицы можно сделать вывод, что в среднем коэффициент корреляции по найденным маркерам выше, чем по маркерам геологов, что позволяет говорить о том, что разработанный программный продукт выделяет границы не хуже экспертов.

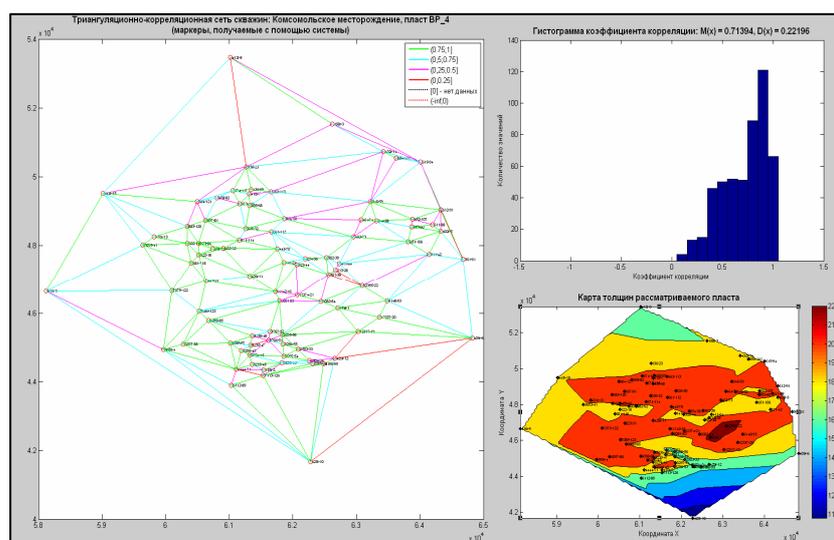


Рис. 4. Пример визуализации результатов, полученных с помощью разработанной системы

Таблица 1

Сравнение границ пластов, полученных с помощью разработанной системы, с границами пластов геологов (Тарасовское месторождение)

Номер скважины	Глубина по стволу скважины до кровли пласта, м		Погрешность		Глубина по стволу скважины до подошвы пласта, м		Погрешность		Мощность пласта, м		Погрешность		Коэффициент корреляции		Погрешность	
	расчет	по оценке эксперта	абсолют, м	относит, %	расчет	по оценке эксперта	абсолют, м	относит, %	расчет	по оценке эксперта	абсолют, м	относит, %	расчет	по оценке эксперта	абсолют	относит, %
Скважина 1247	2664.20	2663.70	0.50	2.81	2682.20	2681.50	0.70	3.93	18.00	17.80	0.20	1.12	0.9183	0.8538	0.0644	7.02
Скважина 1004	2851.40	2850.60	0.80	3.62	2871.40	2872.70	-1.30	-5.88	20.00	22.10	-2.10	-9.50	0.9180	0.8581	0.0599	6.52
Скважина 1020	2650.00	2650.00	0.00	0.00	2668.60	2667.70	0.90	5.08	18.60	17.70	0.90	5.08	0.8337	0.7895	0.0442	5.30
Скважина 1084	2660	2660.3	-0.3	-1.37	2682.2	2682.2	0	0	22.2	21.9	0.3	1.37	0.9518	0.9011	0.0507	5.33
Скважина 1035	2804	2804.8	-0.8	-3.31	2830.2	2829	1.2	4.96	26.2	24.2	2	8.26	0.8101	0.7256	0.0844	10.43
Скважина 3210	2697	2697.9	-0.9	-3.85	2722	2721.3	0.7	2.99	25	23.4	1.6	6.84	0.9441	0.8529	0.0912	9.66
Скважина 3209	2622	2623	-1	-4.33	2646.8	2646.1	0.7	3.03	24.8	23.1	1.7	7.36	0.8989	0.7124	0.1865	20.75
Скважина 3203	2900.4	2901.4	-1	-4.2	2926.2	2925.2	1	4.2	25.8	23.8	2	8.4	0.8834	0.8119	0.0715	8.09
Скважина 1082	2681.2	2682.3	-1.1	-4.68	2706.2	2705.8	0.4	1.7	25	23.5	1.5	6.38	0.9644	0.7677	0.1966	20.40
...

ЗАКЛЮЧЕНИЕ

В статье был предложен подход к решению задачи выделения естественных неструктурированных объектов.

Рассматривался ряд вопросов: от разработки модели до написания программного обеспечения, касающегося выделения продуктивных пластов.

Была сформулирована постановка задачи, которая позволила отнести ее к классу задач оптимизационного геометрического моделирования.

Для решения задачи было применено несколько подходов. Один из них базируется на наиболее распространенном метаэвристическом методе «генетический алгоритм», который был адаптирован для решения поставленной задачи. Кроме того, для решения задачи был использован метод, заимствованный из области распознавания речевых сигналов, – метод динамического искажения времени (DTW).

Все представленные в работе методы, алгоритмы и функции были реализованы в среде MATLAB, которая является удобным средством и с точки зрения встроенных стандартных функ-

ций, так и с точки зрения возможности разработки собственных.

Разработанная система была опробована на Приобском, Тарасовском и Комсомольском месторождениях. В результате проведения вычислительных экспериментов было установлено, что применение разработанного программного продукта позволяет улучшить решение, получаемое экспертами (геологами), в среднем на 5–15% в зависимости от месторождения и выделяемого пласта при этом время работы значительно сокращается.

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян, С. А. Прикладная статистика. Классификация и снижение размерности / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин, В. М. Бухштабер. М. : Финансы и статистика, 1989. 607 с.
2. Байков, В. А. Современные методы анализа геофизических исследований скважин / В. А. Байков, Г. А. Борисов, О. М. Верхотурова, О. В. Надеждин // Науч.-техн. вестник ОАО «НК «Роснефть». 2007. № 1. С. 27–31.
3. Бондаренко, И. Ю. Анализ эффективности метода нечёткого сопоставления образов для распознавания изолированных слов / И. Ю. Бондаренко,

О. И. Федяев // Сб. трудов VI междунар. науч. конф. «Интеллектуальный анализ информации ИАИ-2006». К. : Просвіта, 2006. С. 20–27.

4. **Губерман, Ш. А.** Неформальный анализ данных в геологии и геофизике / Ш. А. Губерман. М. : Недра, 1987.

5. **Губерман, Ш. А.** Корреляция геофизических разрезов скважин на ЭВМ / Ш. А. Губерман, Е.Е. Калинина, М.И. Овчинникова, В.Ф. Осипов // Геология нефти и газа. 1981. № 2.

6. **Гутман, И. С.** Корреляция геолого-геофизических разрезов скважин с помощью программы «Геокор-2» / И. С. Гутман, В. Е. Копылов, Ф. С. Котов, Е. И. Бронскова // Геология нефти и газа. 2002. № 2.

7. **Тенев, В. А.** Интеллектуальные системы интерпретации данных геофизических исследований скважин / В. А. Тенев, Б.А. Якимович, М. А. Семенов, Н. Б. Паклин // Искусственный интеллект. 2002. № 3.

8. **Теория** статистики : учеб. / Под ред. проф. Р.А. Шмойловой. М. : Финансы и статистика, 2002. 560 с.

9. **Herrera, F.** Hybrid crossover operators for real-coded genetic algorithms: an experimental study /

F. Herrera, M. Lozano, A. M. Sanchez // Soft Computing. 2005. Vol. 9, № 4. P. 280–298.

10. **Holland, J. H.** Adaptation in natural and artificial systems / J. H. Holland. The University of Michigan Press, 1975.

ОБ АВТОРЕ



Верхотурова Олеся Михайловна, асп. каф. вычислит. математики и кибернетики УГАТУ. Дипл. экономист-математик (УГАТУ, 2006). Готовит диссертацию по оптимизационному геометрическому моделированию.