

6. Айвазян С.А. Интегральные индикаторы качества жизни населения: их построение и использование в социально-экономическом управлении и межрегиональных сопоставлениях. М.: ЦЭМИ РАН, 2000.

7. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998. 1006 с.

8. Айвазян С.А. Эмпирический анализ синтетических категорий качества жизни. // Экономика и математические методы. 2003. Т. 39. №3. С. 19-53.

9. www.tumstat.gks.ru РОССТАТ (Территориальный орган федеральной службы государственной статистики по Тюменской области — Тюменьстат).

**Артем Валерьевич БЫКОВ** —  
ассистент кафедры автоматизации и вычислительной техники  
artyom@front.ru

**Владимир Эрнестович БОРЗЫХ** —  
зав. кафедрой автоматизации и вычислительной техники,  
доктор физико-математических наук  
borzykh@tsogu.ru

Тюменский государственный нефтегазовый университет

УДК 519.688

## ИСПОЛЬЗОВАНИЕ НЕЧЕТКИХ МНОЖЕСТВ В ГЕОЛОГИИ

### APPLICATION OF FUZZY SETS IN GEOLOGY

**АННОТАЦИЯ.** Рассматривается процесс построения геологических моделей на основе стратиграфической информации с использованием теории нечетких множеств. Такой подход позволяет корректно описывать неопределенности, присущие геологической информации.

**SUMMARY.** The article contains description of geological model development based on stratigraphic data with the implementation of fuzzy sets theory. Such approach allows correct determining of geological data uncertainties.

**КЛЮЧЕВЫЕ СЛОВА.** Нечеткие множества, стратиграфия, геология.

**KEY WORDS.** Fuzzy sets, stratigraphy, geology.

Одним из важных этапов разработки нефтяных и газовых месторождений является создание геологических моделей. При этом в качестве основы таких моделей наряду с традиционными подходами возможно использовать стратиграфическую информацию. Покажем, что это может принести.

Итак, имеется следующая исходная информация:

- данные с пробуренных скважин (их названия, координаты устьев, альтитуды, траектории стволов и др.);
- результаты геофизических исследований скважин;
- информация об исследовании керна;
- разбивки траекторий скважин по стратиграфическим подразделениям (полученные специалистами-геологами на основе других данных);
- метаданные (исполнители — организации, авторы; время получения; использовавшиеся методики и инструменты и др.).

В результате можно получить следующие характеристики для стратиграфических подразделений:

- статистические параметры (средневзвешенное значение, доверительный интервал, медиана и др.) для любого геофизического метода исследования скважин, по которому имеется информация;

- результаты исследования керна; для признаков, измеряемых номинальной шкалой, (тип породы, найденные остатки организмов и др.) — частоты встречаемости; для непрерывных данных (пористость, проницаемость, продуктивность и др.) — получение всех статистических характеристик;
- границы распространения подразделения (используются упоминания подразделений в разбивках скважин и их координаты);
- а также следующую информацию по всем стратонам:
- все встречающиеся подразделения для заданной территории (стратоны должны быть одного ранга и находиться на одной шкале);
- районы распространения различных подразделений одного возраста (например, границы расположения свит, содержащихся в одном горизонте).

Рассмотрим исходные данные. Можно выделить следующие характеристики, которые влияют на их качество:

- существует несколько видов исходных данных, для каждого из которых сложились свои форматы представления, зачастую плохо согласующиеся между собой;
- данные поступали и поступают из различных организаций, форматы представления в которых могут различаться;
- большинство исходного материала было получено в 1960-70-е гг. в период массового бурения скважин, когда использовались неавтоматизированные технологии, переход от которых не был системным (выполнялся разными организациями в течение длительного периода времени, часто без сохранения преемственности), что привело к появлению большого количества некачественных данных;
- во время массового бурения скважин в 1960-70-е гг. качеству получаемых данных не всегда уделялось должное внимание, что было особенно характерно для эксплуатационных скважин.

Кроме того, одной из главных проблем является наличие данных, которые противоречат друг другу. Положение усугубляется тем, что даже эксперты не могут прийти к общей точке зрения, какие из них верные, а какие нет. Такая ситуация случилась из-за наличия небольшого количества достоверных данных, когда геологи вынуждены строить гипотезы, основываясь на своем опыте и интуиции, что приводит к различной интерпретации одних и тех же исходных данных и разделению специалистов на школы и направления. К тому же в рамках одной школы у различных авторов также могут быть расхождения, одни и те же авторы со временем могут менять свое мнение [1, 2].

Все это требует проведения тщательной подготовки исходных данных. Одновременно с очисткой производится их интеграция. Для этого применяются различные методы (как общие: обработка пропущенных данных, дубликатов, выбросов и др. [3], так и специфичные для геологии), кроме того, разрозненная информация приводится к одному виду.

Далее происходит процесс анализа геологических данных. Рассмотрим его более подробно.

Одной из особенностей наук о Земле является отсутствие возможности непосредственно изучить весь массив пород, находящихся глубоко в недрах земли, и точно определить исторические процессы. Поэтому в практической работе приходится полагаться в основном на опыт и интуицию геологов, что часто приводит к субъективным результатам.

Приемлемым выходом из сложившейся ситуации является использование теории нечетких множеств [3]. Построенные на ее основе модели позволяют

уменьшить требования к исходным данным и корректно описывать в них неопределенности.

Алгоритм подготовки исходной информации следующий. По имеющимся разбивкам скважин, координатам скважин, обобщающим геологическим коэффициентам, а также по другим характеристикам строится матрица степени принадлежности стратиграфических подразделений скважинам, которая определяет, насколько конкретные участки скважины соответствуют подразделениям. Рассмотрим способы построения такой матрицы.

Для этого обратимся сначала к алгоритмам районирования распространения стратиграфических подразделений на возрастных срезах. Построенные с их помощью карты литолого-стратиграфического районирования территории играют важную роль на региональных стратиграфических схемах. На этих картах изображаются географические границы распространения региональных и местных стратиграфических подразделений.

Стратоны с большей площадью распространения более привычны и чаще используются геологами, даже там, где в этом нет необходимости. Границы же не могут быть точно определены из-за естественной изменчивости свойств горных пород. Таким образом, способ жесткого разграничения стратиграфических подразделений не всегда приемлем и может давать результаты, которые отличаются от данных разбивок скважин.

Альтернативным методом разграничения стратиграфических подразделений является использование нечетких переменных. В этом случае каждая точка принадлежит всем стратонам (степень принадлежности от 0 до 1). Это позволяет корректно обрабатывать граничные территории, которые одновременно принадлежат разным подразделениям. При удалении от центра расположения одного стратона к центру другого степень принадлежности к первому и второму уменьшается и растет соответственно.

На полученной таким образом карте можно выделить области с высокой степенью принадлежностью для определенного подразделения, например, от 90%, и считать их областями распространения данного стратона. Кроме того, если на карту нанести изолинии степеней принадлежности стратиграфических подразделений, то возможно визуально контролировать распространение этих подразделений на карте, а также переходные зоны.

Наиболее популярным является алгоритм анализа частот употребления стратонов при проведении разбивок [4].

В этом случае для каждой точки на карте вычисляется следующее соотношение для каждой свиты конкретного возрастного среза:

$$P = S / N,$$

где  $p$  — степень принадлежности породы свите,  $S$  — количество упоминаний свиты в окрестностях данной точки,  $N$  — общее количество упоминаний всех свит в окрестностях данной точки.

Если количество данных в окрестностях ниже минимально допустимого уровня, то степень принадлежности не вычисляется. Так как для одной скважины может существовать несколько разбивок разных авторов (возможно, с различными именами свит), то учитывается именно количество упоминаний свит, а не скважин.

Однако можно предложить другой метод, основанный на анализе свойств стратонов. Так как разбивки подразделений выполняются по данным, полученным при различных видах исследования пород, в первую очередь из скважин,

то, если взять исходные данные и разбивки, можно найти признаки, по которым производились разбивки. Эти признаки должны максимально присутствовать в центре стратона и проявляться в меньшей степени у окраин, где начинают появляться признаки соседнего подразделения. Используя эти признаки, можно составить формальные оценки, которые говорят о степени принадлежности к тому или иному стратону.

Основная задача состоит в следующем. Для каждого набора характеристик необходимо определить степени принадлежности всем подразделениям. Представим это в формальном виде.

Дано множество объектов:

$$I = \{i_1, i_2, \dots, i_j, \dots, i_n\},$$

где  $i_j$  — информация по скважине.

Каждый объект характеризуется множеством переменных:

$$i_j = \{x_1, x_2, \dots, x_h, \dots, x_k\},$$

где  $x_h$  — независимые переменные, значения которых известны.

Каждая переменная  $x_h$  может принимать значения из некоторого множества:

$$R_h = \{r_{h1}, r_{h2}, \dots\}.$$

Данное множество может быть конечным или бесконечным.

На основании этих переменных выбирается значение матрицы принадлежности  $U$ . Значение  $u_{ij}$  показывает степень принадлежности скважины  $i$  подразделению  $m$ , причем должно выполняться следующее условие:

$$u_{ij} \in [0, 1].$$

Следует отметить, что сумма любой строки будет больше 1. Это связано с тем, что каждая скважина содержит стратиграфические подразделения разных возрастов, которые располагаются на разной глубине. Если же матрицу  $U$  разбить на подматрицы, в каждой из которых будут только стратоны одного возраста, то в этом случае сумма каждой строки подматрицы будет равна 1.

Кроме того, здесь рассматриваются подразделения с малой мощностью, на которые обычно разбивают траектории скважин. Степени принадлежности стратонов, стоящих выше в иерархии, получаются из объединения степеней принадлежности входящих в них подразделений.

Рассмотрим основные алгоритмы, которые используются для решения подобных задач. Это алгоритмы классификации или кластеризации. В случае классификации имеется обучающее множество объектов, для которых известны значения переменных и класс. Необходимо каждый поступающий новый объект по известным значениям переменных отнести к одному из классов. Кластеризация отличается тем, что обучающее множество отсутствует, необходимо разбить объекты по известным значениям переменных на кластеры (группы) похожих объектов.

Для данной задачи классы объектов (принадлежность к определенному стратону) известны, однако эта информация является дополнительной. В основе нового разделения на классы лежит анализ свойств пород. В результате

строится новая матрица принадлежности, которая может по некоторым параметрам существенно отличаться от первоначального разделения объектов.

Рассмотрим алгоритмы кластеризации [3, 5]. В них выделяют иерархические и неиерархические. Среди иерархических нет алгоритмов, которые позволяют работать с нечеткими множествами, поэтому далее мы их касаться не будем. Среди неиерархических наиболее популярными является метод нечеткой самоорганизации *c-means* и его обобщение в виде алгоритма Густафсона-Кесселя, которые находят множества с ядром в центре кластера. Однако существуют и алгоритмы другого вида, которые обходятся без использования центра кластера. Хотя такие алгоритмы используют нечеткие отношения и носят более универсальный характер, их использование проблематично из-за того, что они дают в результате множество разбиения, а не степени принадлежности.

Существует большое количество алгоритмов классификации [3, 5]. Однако большинство из них дают дискретные результаты. Степени принадлежности дают только алгоритмы, создающие правила классификации на основе статистики. Это относительно простой *1R*-алгоритм и алгоритм байесовской классификации. Одним из недостатков этих алгоритмов является невозможность прямой обработки непрерывных переменных, их необходимо преобразовывать к дискретному виду, что может привести к потере закономерностей.

Таким образом, имеются алгоритмы классификации и кластеризации (будем использовать алгоритмы Густафсона-Кесселя и байесовской классификации, как наиболее совершенные в своих группах), необходимо найти среди них оптимальный. Для этого дополним их для решения данной задачи.

В случае классификации исходное множество данных интервалов скважин, на основе которого первоначально создаются правила, затем прогоняется через эти правила для определения степени принадлежности стратонам. В этом случае необходимо следить, чтобы не происходило переобучения [5]. Под переобучением понимается слишком точное соответствие правил конкретному обучающему множеству (то есть исходным разбиениям), когда теряется способность к обобщению. В нашем случае это приведет к тому, что результаты будут зависеть больше от разбиений конкретных авторов, а не от непосредственных свойств подразделений. В худшем случае мы получим исходные данные.

В случае кластеризации появляется другая проблема. Так как алгоритм самостоятельно делит все множество информации по разбиениям на группы, то получившиеся группы могут даже приблизительно не совпадать с данными, полученными от авторов разбиений. Это еще связано с тем, что в своей работе геологи полагаются и на плохо формализуемые характеристики, которые невозможно учесть при автоматической обработке.

Если же данные совпадают, то это говорит о хорошем выборе характеристик для разделения стратонов и высокой степени использования всей имеющейся информации в автоматическом режиме.

Однако добиться такого совпадения при кластеризации чрезвычайно сложно, поэтому остановимся на байесовской классификации, которая приводит к значениям, близким к авторским разбиениям, поскольку на них опирается.

Кроме того, результаты, полученные при использовании анализа частот употребления стратонов и при классификации, могут существенно отличаться, хотя это не является достаточным основанием, чтобы говорить о неверных вычислениях. К тому же нельзя напрямую сравнивать степени принадлежности, полученные разными методами. При анализе употребления частот значения степеней принадлежности будут существенно выше, чем при классификации, здесь скорее следует сравнивать их отношения.

Непосредственно из полученной матрицы принадлежности строятся карты распространения стратиграфических подразделений с использованием координат скважин и интерполяцией данных в промежутках между скважинами. Так же, как и при выполнении классификации, необходимо учитывать иерархию подразделений, так как разбивки выполняются для подразделений, занимающих нижние места в иерархии, а карты, как правило, строятся для более крупных подразделений, включающих несколько стратонов более низкого ранга — то есть должно происходить объединение данных подчиненных подразделений низшего ранга.

Покажем еще одно преимущество использования нечетких множеств.

При использовании статистических методов для обработки геологической информации данные, полученные из различных источников, могут различаться по степени достоверности, точности.

Рассмотрим поисково-разведочные и эксплуатационные скважины. Как показывает практика, достоверность данных, полученных с первых, выше, чем со вторых. Часто даже данные с эксплуатационных скважин не берутся в расчет из-за их низкого качества.

Таким образом, существует ряд признаков, которые позволяют судить о качестве исходных данных и их совместимости между собой. Это дает возможность использовать следующий метод повышения достоверности статистической информации. Каждой переменной, которая влияет на качество исходной информации, ставится в соответствие определенный вес. Если производить вычисление статистических характеристик с этими весами, то будет учитываться качество исходных данных и их совместимость друг с другом.

#### *Информация по стратиграфическим подразделениям*

При вычислении характеристик стратиграфических подразделений необходимо использовать максимальное количество доступной информации. Однако в этом случае на границах подразделений (где одни стратоны плавно перетекают в другие) будут учитываться и свойства пород соседних подразделений. Чтобы уменьшить влияние граничных районов, вводятся веса, которые уменьшаются от центра к периферии.

В этом случае целесообразно в качестве весов использовать степени соответствия стратоном.

#### *Информация по конкретным координатам*

В этом случае необходимо получить информацию по конкретному месту, в котором часто непосредственно нет скважин. То есть необходимо взять данные из близлежащих скважин и провести их интерполяцию с учетом расположения.

Задачи такого рода часто встречаются при работе с географическими информационными системами и хорошо проработаны [6]. Рассмотрим только те

методы, которые позволяют работать с точками (у нас — скважины), расположенными через разные расстояния.

Так как разные скважины могут иметь разное расстояние от интересующего нас места, то в качестве одного из компонента весов необходимо использовать величину, обратно пропорциональную квадрату расстояния до скважины. Данный метод получил название обратных взвешенных расстояний.

Рассмотрим рис. 1, на нем приведен пример, когда скважина 1 должна иметь больший вес, чем любая другая скважина, так как с этой стороны нет других скважин, а скважины 2, 3 и 4 расположены близко друг к другу. Для урегулирования таких ситуаций необходимо ввести другой компонент веса, а именно — величину, обратную сумме квадратов расстояний до всех остальных исходных скважин.

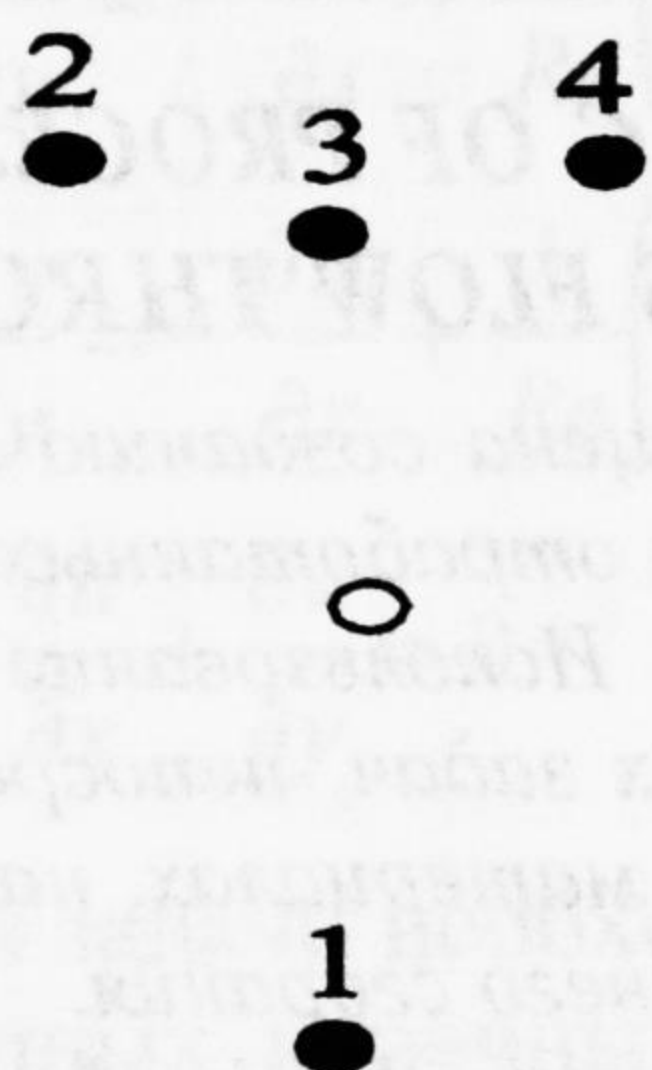


Рис. 1. Пример, когда расположение скважины влияет на ее вес

Таким образом, вся информация будет получена из среднего значения всех скважин с учетом весов. О встречающихся в данном месте стратиграфических подразделений можно узнать из степеней соответствия, которые определяются для каждой скважины.

Итак, основываясь на матрице степеней принадлежности, можно вычислить статистические характеристики по конкретным подразделениям или районам.

Полученные характеристики и районы распространения используются при исследовании малоизученных территорий, а также для оценки общей ситуации по стратиграфическим подразделениям.

#### СПИСОК ЛИТЕРАТУРЫ

1. Филиппович Ю.В. // Вестник недропользователя Ханты-Мансийского автономного округа. 2002. № 8. С. 40-47.
2. Киричкова А.И., Чижова В.А., Сташкова Э.К. // Нефтегазовая геология. Теория и практика: электр. науч. журн. 2007. № 2. 32 с.
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2004. 336 с.
4. Каталог литолого-стратиграфических разбивок разрезов поисково-разведочных скважин. Том 1. Ханты-Мансийский автономный округ / Под ред. В.Ф. Гришкевича, Е.А. Теплякова. Ханты-Мансийск: Путиведь, 2000. С. 419-422.
5. Чубукова И.А. Data Mining. М.: БИНОМ. Лаборатория знаний, 2008. 382 с.
6. ДеМерс М.Н. Географические информационные системы. Основы: Пер. с англ. М.: Дата+, 1999. С. 295-310.