

А. В. Быков

ИСПОЛЬЗОВАНИЕ НЕЧЕТКИХ МНОЖЕСТВ ПРИ РЕШЕНИИ ЗАДАЧ СТРАТИГРАФИИ

При разработке углеводородных месторождений обращаются к созданию геологических моделей. Оправданным в этом случае является использование информации о стратиграфических подразделениях: статистические параметры для геофизических методов, результаты исследования керна, границы распространения подразделения. Подготовленные данные анализируются с помощью теории нечетких множеств, что позволяет «смягчить» требования к источникам информации и корректно описывать неопределенности.

Геологическая модель, стратиграфические параметры, месторождение, нефть, газ.

Для решения вопросов прогнозирования в процессе разработки нефтяных и газовых месторождений обращаются к созданию геологических моделей. Оправданным в этом случае является использование стратиграфической информации, и не только для данного объекта, но и для подобных стратиграфических подразделений в других районах.

В работе со стратиграфическими подразделениями используют:

- информацию о пробуренных скважинах (названия, координаты устьев, альтитуды, траектории стволов и др.);
- результаты геофизических исследований скважин;
- данные исследования керна;
- сведения о разбивке траекторий скважин на стратиграфические подразделения (полученные специалистами на основе других данных);
- метаданные (исполнители — организации, авторы; время получения; использовавшиеся методики и инструменты; ссылки на исходные данные и др.).

Основываясь на этой информации, необходимо получить следующие характеристики для стратиграфического подразделения:

- статистические параметры (средневзвешенное значение, доверительный интервал, медиана и др.) для любого геофизического метода, по которому имеется информация;
- все 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.) — все статистические характеристики;
- границы распространения подразделения (используются упоминания подразделений в разбивках скважин и их координаты).

Кроме того, необходимо отображать обобщающую 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).

Так как все исходные данные поступают из разнообразных источников, имеют разную степень точности, достоверности, частично дополняют, а порой и противоречат друг другу (не всегда в явном виде), то необходимо провести тщательную подготовку данных.

Одновременно с очисткой данных производится их интеграция. Применяются различные методы очистки, как общие [1], так и специфические для геологии; кроме того, разрозненная информация приводится к единому виду.

Далее организуется процесс анализа геологических данных. Рассмотрим его более подробно, с учетом особенностей, присущих геологии.

Науки о Земле занимают особое место среди других наук. С одной стороны, геология опирается на данные и методы точных наук, с другой — из-за невозможности непосредственно изучить весь массив пород, залегающих глубоко в недрах, и определить исторические процессы приходится в практической работе полагаться в основном на опыт и интуицию экспертов, что приводит к субъективным результатам.

Использование математического аппарата обоснованно в первую очередь для хорошо изученных процессов (например, физические, которые можно смоделировать на земной поверхности). Это не всегда приемлемо для наук о Земле.

Выходом из сложившейся ситуации является использование теории нечетких множеств [1]. Построенные на ее основе модели позволяют «смягчить» требования к исходным данным и дают возможность корректно описывать неопределенности.

Алгоритм подготовки исходной информации следующий: по имеющимся данным разбивки, координатам скважин, обобщающим геологическим коэффициентам и другим характеристикам строится матрица степени принадлежности стратиграфических подразделений скважинам, которая позволяет определить, насколько конкретные участки скважины соответствуют подразделениям. Рассмотрим способы построения такой матрицы.

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

При разграничении распространения стратиграфических подразделений возникают следующие трудности. Стратоны с большей площадью распространения более привычны и чаще используются геологами, даже если 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 со степенью 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 [3].

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 возрастном срезе в окрестностях данной точки;

N — общее количество упоминаний всех свит возрастного среза 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 краин, где начинают появляться признаки соседнего подразделения. Используя эти признаки, можно составить формальные оценки, которые говорят о степени принадлежности к тому или иному стратону.

В этом случае необходимо использовать алгоритм, который автоматически находит эти признаки и определяет степень принадлежности к тому или иному подразделению. Одной из необходимых особенностей алгоритма должен быть вывод результатов в виде нечетких множеств.

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_{nm} показывает степень принадлежности скважины n 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 (к какому стратону относятся) известны, однако эта информация является дополнительной, а за основу нового разделения на классы принимаются свойства пород. В результате строится новая матрица принадлежности, которая по некоторым параметрам может существенно отличаться от первоначального разделения объектов.

Рассмотрим алгоритмы кластеризации [1, 6]. Среди 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 являются метод нечеткой самоорганизации *c-means* и его обобщение в виде алгоритма Густафсона — Кесселя, которые находят множества с ядром в центре кластера. Однако есть и алгоритмы другого вида, которые обходятся без использования центра кластера. Хотя такие алгоритмы используют нечеткие отношения и носят более универсальный характер, их использование проблематично, так как они дают в результате множество разбиения, а не степени принадлежности.

Существует большое количество алгоритмов классификации [1, 6]. 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 в дискретный вид, 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 выбрать оптимальный. Дополним их для решения данной задачи.

В случае классификации исходное множество данных интервалов скважин, на основе которого первоначально создаются правила, затем пропускается через эти правила для определения степени принадлежности стратонам. При этом необходимо следить, чтобы не происходило переобучения [6]. Под переобучением понимается слишком точное соответствие правил конкретному обучающему множеству (т. е. исходным разбивкам), когда теряется способность к обобщению. В данных условиях это приведет к тому, что результаты будут зависеть в большей степени от разбивок конкретных авторов, чем от непосредственных свойств подразделений. В худшем случае мы получим исходные данные. С другой стороны, должны быть выработаны такие правила, чтобы классификация приводила к результатам, достаточно близким к исходным разбивкам скважин.

В случае кластеризации появляется другая проблема. Так как алгоритм самостоятельно делит все множество информации по разбивкам на группы, то получившиеся группы могут даже приблизительно не совпадать с данными, полученными от авторов разбивок. Это связано еще 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 разбивкам, так как на них опирается.

Следует отметить, что в данном случае мы приравниваем вероятность принадлежности подразделению к степени соответствия ему. Это объясняется тем, что последняя является более широким понятием, которое не обязательно должно быть нормализованным (т. е. сумма всех степеней принадлежности не обязательно равна 1).

Кроме того, результаты, полученные при использовании анализа частот употребления стратонов и при классификации, могут существенно различаться, 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: 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 разных источников, могут различаться по степени достоверности, точности. Кроме того, существуют школы, которые одни и те же исходные данные могут интерпретировать по-разному, что не позволяет совместно использовать полученную от них информацию [4, 5].

Рассмотрим поисково-разведочные и эксплуатационные скважины. Как показывает практика, достоверность данных с первых выше, чем со вторых. Часто данные с эксплуатационных скважин даже не берутся в расчет из-за их низкого качества. Основные причины этого перечислены ниже (табл.).

**Основные причины различия качества данных
с поисково-разведочных и эксплуатационных скважин**

Характеристика	Поисково-разведочные скважины	Эксплуатационные скважины	Примечание
Основное назначение скважин	Исследование недр	Добыча полезных ископаемых	
Траектория скважин	Вертикальная	Горизонтальная	Данные инклинометрии могут быть не очень точными
Степень внимания геологов (изученности)	Высокая	Низкая	
Качество отбора материала	Высокое	Низкое	

Таким образом, существует ряд признаков, которые позволяют судить о качестве исходных данных и их совместимости. 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.

Следует отметить, что значение каждого признака может меняться в зависимости от задачи. При наличии двух (или более) школ целесообразно рассчитать статистические характеристики для каждой школы, т. е. выставлять каждый раз наиболее высокие веса представителям данной школы и наиболее низкие ее противникам.

Покажем, как на основе матрицы степени принадлежности можно получить интересующую нас информацию.

Информация по стратиграфическим подразделениям

При вычислении характеристик стратиграфических подразделений необходимо использовать максимальное количество доступной информации. Однако в этом случае на границах подразделений (где одни стратона 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 ответственности стратонам.

Информация по конкретным координатам

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

Задачи такого рода часто встречаются при работе с географическими информационными системами и хорошо проработаны [2]. Рассмотрим только те методы, которые позволяют работать с точками (в данном случае — скважинами), расположенными на разных расстояниях.

Так как скважины могут находиться на неодинаковом расстоянии от интересующего нас места, то в качестве одного из компонентов весов необходимо

использовать величину, обратно пропорциональную квадрату расстояния до скважины. Данный метод получил название метода обратных взвешенных расстояний. Существуют его модификации, учитывающие барьеры, через которые не могут распространять свое влияние соседние точки [7]. Однако в нашем случае наличие таких объектов (например, скал) маловероятно, поэтому оставим такие модификации без внимания.

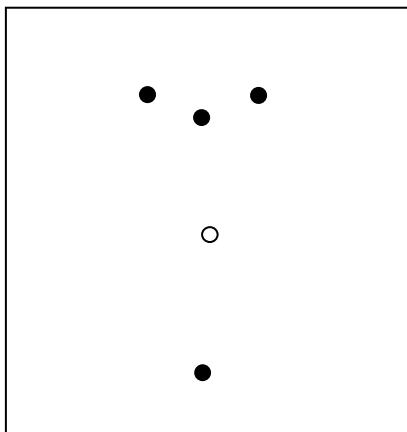


Рис. 1. Пример ситуации, в которой расположение скважины влияет на ее вес

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

Кроме того, необходимо определить максимальное расстояние, на котором будут учитываться скважины. Если это расстояние чрезмерно большое, то будут учитываться лишние объекты, что увеличивает вычислительную нагрузку за счет скважин, которые практически не влияют на конечный результат. При очень маленьком расстоянии важные данные многих объектов не будут учтены, возможно необоснованное появление «белых пятен» (территорий, на которых будут отсутствовать какие-либо результаты).

Стандартным методом оценки такого расстояния является использование вариограммы, на которой по горизонтальной оси откладываются расстояния между скважинами, а по вертикальной — полудисперсия (половина квадрата стандартного отклонения между значениями на этих скважинах). На построенном таким образом графике можно увидеть на начальном этапе рост, который после значения, называемого предельным радиусом корреляции, прекращается (рис. 2). Значение предельного радиуса корреляции и будет требуемым расстоянием. Однако данный метод теряет свою привлекательность, становится малоэффективным и его невозможно формализовать при наличии большого уровня локального шума, т. е. в нашем случае. По этой же причине мы не стали использовать для интерполяции метод кригинга, который также основывается на вариограммах.

Исходя из [3], а также из фактического расположения скважин на территории Ханты-Мансийского автономного округа, возьмем в качестве базового значения расстояние в 25 км. Это значение впоследствии можно будет изменять в зависимости от района и практических результатов использования метода.

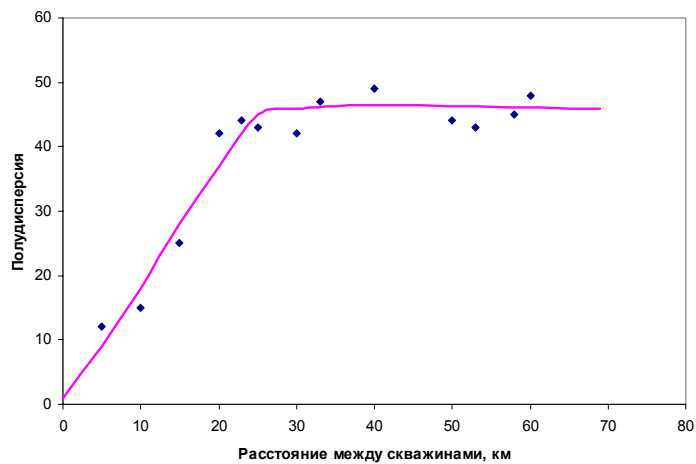


Рис. 2. Пример вариограммы

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

Информация по территории

Рассмотрим построение карт распространения стратиграфических подразделений. Для этого будем использовать информацию, полученную для всех точек. В этом случае на карте можно показать ограниченное число параметров; чтобы показать все параметры, нужно построить несколько карт.

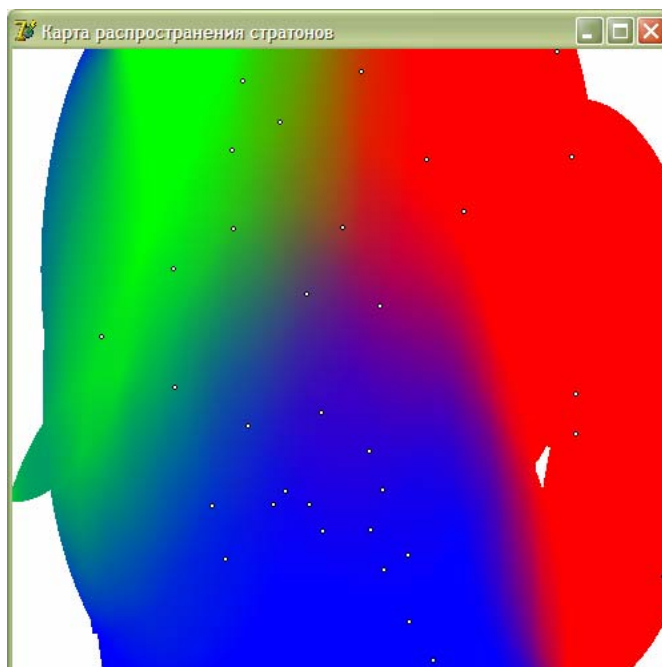


Рис. 3. Пример карты распространения стратонов

Пример карты, созданной на основе тестовых данных, представлен на рис. 3. Здесь точками показаны скважины, белые области говорят об отсутствии достаточной информации (слишком мало близлежащих скважин), а цвет — о подразделении (три основных стратона).

Итак, используя матрицу степеней принадлежности в качестве весов, можно вычислить статистические характеристики по конкретным подразделениям или районам. Полученные характеристики используются при исследовании малоизученных территорий, а также для оценки общей ситуации распространения и свойств стратиграфических подразделений.

ЛИТЕРАТУРА

1. 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.
2. Де Мерс М. Н. Географические информационные системы: Основы: Пер. с англ. — М.: Дата+, 1999. — С. 295–310.
3. Каталог литолого-стратиграфических разбивок разрезов поисково-разведочных скважин. Т. 1: Ханты-Мансийский автономный округ / Под ред. В. Ф. Гришкевича, Е. А. Теплякова. — Ханты-Мансийск, 2000. — С. 419–422.
4. Киричкова А. И., Чижова В. А., Сташкова Э. К., Фортунатова Н. К., Шурыгин Б. Н. Стратиграфия в нефтяной геологии: методология исследований и актуальные проблемы [Электрон. ресурс] // Нефтегазовая геология. Теория и практика. — Электрон. науч. журн. — 2007. — № 2.
5. Филиппович Ю. В. Некоторые аспекты стратиграфического расчленения мезозоя Западной Сибири // Вестн. недропользователя Ханты-Мансийского автономного округа. — 2002. — № 8. — С. 32–40.
6. Чубукова И. А. Data Mining. — М.: БИНОМ. Лаборатория знаний, 2008. — 382 с.
7. Shepard D. A Two-dimensional Interpolation Function for Irregularly Spaced Data // In Processing of Twenty-third National Conference of the Association for Computing Machinery. — 1968. — P. 517–524.

A. V. Bykov

USE OF FUZZY SETS UNDER SOLVING STRATIGRAPHY PROBLEMS

Under hydrocarbon development, they normally turn to geological modeling. Justified here, being use of stratigraphic data. In this case, it is necessary to obtain information on stratigraphic units, i.e. statistical characteristics for geophysical methods, test results on core samples, as well as expansion limit of these units. The prepared data is investigated using theory of fuzzy sets, which enables "to soften" requirements to data sources, correctly describing uncertainties.

Geological model, stratigraphic characteristics, field, oil, gas.