

Николаева А.Н.,

Студент

4 курс, Отделение геологии

Национальный исследовательский Томский политехнический

университет

Россия, г. Томск

АНАЛИЗ ФОРМИРОВАНИЯ БАЗЫ ГЕОЛОГИЧЕСКИХ ЗНАНИЙ

***Аннотация:** В геологии актуализировались ГИС, внедряются экспертные системы, технологии эволюционного роста. Разрабатываются сложные программные комплексы для работы с разнообразной информацией. В работе проведен системный анализ задачи разработки экспертных систем, баз знаний геологической направленности. Предложена процедура вывода знаний.*

***Ключевые слова:** нефтегазовая отрасль, геологоразведка, самоорганизация, знания, вывод, база.*

***Annotation:** GIS has been updated in geology, expert systems, and technologies for evolutionary growth are being introduced. Complex software systems are being developed for working with a variety of information. A system analysis of the task of developing expert systems and geological knowledge bases has been carried out in the work. The procedure for the withdrawal of knowledge is proposed.*

***Keywords:** oil and gas industry, exploration, self-organization, knowledge, conclusion, base.*

Плохо формализуемые (структурируемые) системы, как правило, сложны не только с точки зрения системного анализа [1], но и с точки зрения выбора релевантного инструментария аналитика, их применения при исследовании. Одним из таких инструментов являются, например, нечёткие множества и

логика, эволюционное моделирование и генетические алгоритмы, нейросистемы и сети.

Повысить эффективность и качество решений, геологоразведки можно актуализацией ИТ-ориентированных интегрированных систем принятия решений (ИСППР), позволяющих оперативно добывать информацию. Например, DataMining, системы ERP (в динамическом производственном процессе) или Enterprise-Wide Risk Management (в комплексном риск-менеджменте предприятия).

Используются не только геологические БД, но и БЗ (Базы Знаний), в частности, с 1998 года «БЗ: гидрогеология, инженерная геология, геоэкология» модульной структуры (Библиотека, Калькулятор, Карты и др.). Это ядро АРМ инженера-геолога.

Основными препятствиями к использованию этих и других инструментов стали:

- 1) субъективизм, неопределённость, неполнота используемых экспертных данных, отсутствие релевантных эвристических процедур;
- 2) сложность идентификации (параметризации) функций принадлежности;
- 3) отсутствие (сложность) адаптационных, самоорганизационных механизмов баз знаний (правил вывода из них).

Постановка, формализация проблемы. Требуется идентифицировать по статистическим данным функцию:

$$y = f(x_1, x_2, \dots, x_n) \equiv f(x), \quad x = (x_1, x_2, \dots, x_n)$$

используя систему продукции (правил вывода) $P = \{p_i: i = 1, 2, \dots, m\}$,

$$p_i: \text{if} \left(\bigwedge_{j=1}^m (x_j = a_{ij}) \right) \text{ then } (y = b_i)$$

где a_{ij}, b_i – элементы нечетких множеств, $A = \{a_{ij}: i = 1, 2, \dots, m; j = 1, 2, \dots, n\}$, $B = \{b_j: j = 1, 2, \dots, m\}$, соответственно, с функциями принадлежности [2] $\gamma_{ij}(x_j), \mu(y)$.

Правило P_i означает: «если все x_i близки к a_{ij} , то y – также близок к b_i ». Вид, количество m правил вывода (продукций) должны быть определены исходя из условий обеспечения критерия адекватности прогнозных (теоретических) $Y^T(x^{(k)})$ и статистических (экспериментальных) $y^{(k)}$ значений:

$$F(a) = \frac{1}{N} \sum_{k=1}^N (y^T(x^{(k)}, a) - y^{(k)})^2 \Rightarrow \min,$$

где a – вектор идентифицируемых параметров функций принадлежности, N – количество экспериментальных данных.

Для самоорганизации базы продукций (правил вывода) можно использовать процедуру, аналогичную [3].

Считаем, вид функции принадлежности параметрически определен, например, функцией типа Гаусса:

$$\frac{(x - c)^2}{2\delta^2}$$

где c, δ – параметры центра и размах относительно неё (см. рис.1).

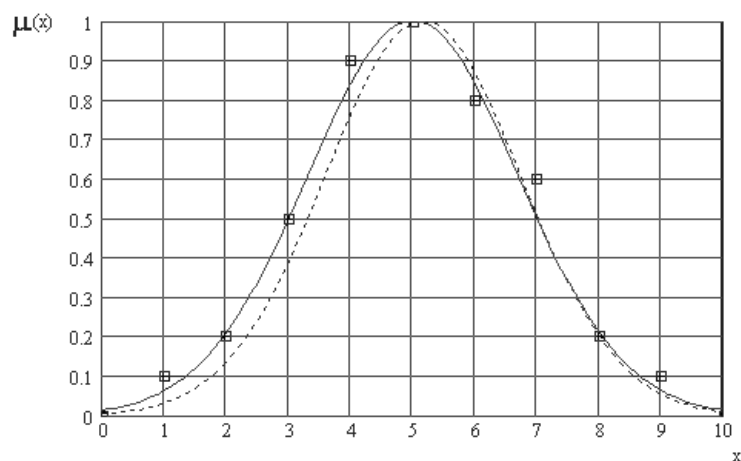


Рис.1. Аппроксимация дискретного ряда (точки)т непрерывной Гауссовой функцией принадлежности (пунктирная линия)

Процедура состоит из этапов.

1. Предварительный анализ статистических данных – выбраковка грубых данных, нахождение основных статистических значений x_j^{min} , x_j^{max} , x_j^{opt} , \tilde{x}_j –

минимальных, максимальных, оптимальных и шкалированных (нормированных) переменных, где

$$\tilde{x}_j = \frac{|x_j - x_j^{opt}|}{x_j^{max} - x_j^{min}}.$$

2. Выбираем первую обучающую выборку: $k := 1$.

3. Формируем вектор $\langle x^{(1)}, y^{(1)} \rangle$ для начальной базы правил P из одного правила P_1 , у которых функции принадлежности $\gamma_{1j}(\tilde{x}_j)$ и $\mu_1(\tilde{y})$ имеют центры $\tilde{x}_j^{(1)}, \tilde{y}^{(1)}$ и отклонения $\delta_0^{(1)} = \delta_y = 0.5, j = 1, 2, \dots, n$

4. Выбираем первую продукцию: $m := 1$.

5. Пока $k \leq N$ выполняем нижеследующие пункты:

5.1. Предъявляем обучающую выборку $\langle x^{(k)}, y^{(k)} \rangle$.

5.2. Находим оценку (согласно правилам P) $\tilde{y}^{(k)}(x^{(k)})$.

5.3. Если $|\tilde{y}^{(k)} - y^{(k)}| > \varepsilon$, где ε – задаваемая точность (критериальная мера) обученности, $k := k + 1$ то $m := m + 1$.

5.4. Вводим новое правило P_i с центрами $\tilde{x}_j^{(k)}$ и $\tilde{y}^{(k)}$, отклонениями $\delta_j^{(k)} = \delta_0 / \sqrt[n+1]{m}$, это гарантирует (см. [4]) сходимость $\{\tilde{f}(x)\} \rightarrow f(x)$ при $N \rightarrow \infty$.

5.5. Переходим к следующей обучающей выборке: $k := k + 1$.

Приведённая процедура не связана с ограничениями по объему выборки $y^{(k)}$, она достаточно просто реализуема при имитационном моделировании. В качестве практического механизма нечеткого вывода можно взять процедуру Мамдани [5].

При моделировании приходится разрабатывать сложные, большие программные комплексы вывода знаний. Они должны удовлетворять следующим требованиям:

- 1) уметь работать с разнообразной информацией;
- 2) иметь несложный, комфортный интерфейс, близкий к обычно используемому на предприятиях;
- 3) учитывать промышленные и отраслевые нормы и ограничения;

4) защищать от несанкционированного доступа к данным.

Заключение. Самоорганизация базы продукционных правил нечетко-ориентированных СНН или СППР, а так же систем ситуационного (имитационного) моделирования – необходимые условия релевантного, содержательного принятия решений, выводов о причинно-следственных связях факторов и отклика геологической системы. Важно иметь вычислительно гибкие, с компактной системой продукций, процедуры самоорганизации, базирующиеся на релевантных функциях принадлежности, прозрачной (даже «интуитивно понятной») лингвистической интерпретацией типа «очень малый», «очень точный» и др.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Казиев В.М. Введение в анализ, синтез и моделирование систем – М.: Бинوم. Лаборатория знаний. Интуит.ру. -2007. -244с.
2. Алтунин А.Е., Семухин М.В. Модели и алгоритмы принятия решений в нечетких условиях. –М.: Изд-во Тюменского госуниверситета, 2000. –352с.
3. Круглов В.В., Усков А.А. Два подхода к самоорганизации базы правил системы нечеткого логического вывода // Информационные технологии, №2, 2006, с.14-18.
4. Катковник В.Я. Непараметрическая идентификация и сглаживание данных: метод локальной аппроксимации. –М.: Наука.-1985. -249с.
5. Леоненков В.В. Нечеткое моделирование в среде MatLab и fuzzyTECH. –СПб.: БХВ-Петербург. -2003. -736с.