

## О КЛАССИФИКАЦИИ ГЕОЛОГИЧЕСКИХ ДАННЫХ МЕТОДАМИ МЯГКИХ ВЫЧИСЛЕНИЙ

И.З. Батыршин, А.Х. Салимов

Методы теории мягких вычислений, включающие в себя нечеткие модели и нейронные сети, широко применяются в задачах классификации данных [1-6]. В работе исследуются методы классификации пористости пород на основе данных комплекса измерений параметров пород по скважине [7, 8]. Определение пористости пород используется для оценки перспективности скважин в нефтедобыче. Рассматривается метод классификации на основе нечеткого гранулирования классов и метод классификации с помощью нейронных сетей.

В работе [7] ставилась задача определения минимального набора признаков для классификации пористости пород с целью минимизации затрат на необходимый комплекс измерений и оценки возможности использования этих методов для пересмотра перспективности законсервированных скважин, по которым имеется информация лишь по небольшому набору параметров. В [7] на основе применения статистических методов и анализа наличия информации по старым скважинам из набора в более чем 15 признаков был выделен минимальный набор из пяти признаков для решения задачи классификации пористости пород. Этими признаками являются:  $p_1$  - DT4P (Delta-T Comp-P&S),  $p_2$  - RHOV (Bulk Density),  $p_3$  - NPHI (Thermal Neutron Porosity),  $p_4$  - GR (Gamma Ray),  $p_5$  - MSFL (MSFL Resistivity). В данной работе на основе информации об описании пород набором векторов значений по этим пяти признакам и известной классификации пород на три класса пористости:  $C_1$  (пористо-кавернозный),  $C_2$  (первичный),  $C_3$  (пористо-трещиноватый), решалась задача построения классификаторов на основе нечетких гранул и нейронных сетей. Методы построения классификаторов и результаты их применения для классификации пород описываются далее.

### Классификация с помощью нечетких гранул

В данном методе три класса пористости  $C_1$ ,  $C_2$  и  $C_3$  представляются в виде нечетких подмножеств  $G_1$ ,  $G_2$ ,  $G_3$  в пятимерном пространстве признаков следующим образом. Каждый нечеткий класс  $G_i$  ( $i=1,2,3$ ) опреде-

ляется как объединение  $N_i$  нечетких гранул  $G_i = G_{i1} \cup G_{i2} \cup \dots \cup G_{ij} \cup \dots \cup G_{iN_i}$ , где нечеткие гранулы  $G_{ij}$  ( $j=1, \dots, N_i$ ) получаются в результате пересечения соответствующих функций принадлежности  $\mu_{ijk}$ , определенных на области значений параметров  $p_k$  ( $k=1, \dots, 5$ ). В качестве таких функций принадлежности использовались обобщенные колоколообразные функции принадлежности [4]:

$$\mu_{ijk}(x) = \frac{1}{1 + \left( \frac{x - c_{ijk}}{a_{ijk}} \right)^{2b_{ijk}}},$$

каждая из которых определялась тремя параметрами ( $a, b, c$ ). Общее число параметров всех нечетких множеств равно  $(N_1 + N_2 + N_3) \cdot 5 \cdot 3 = 15 \cdot (N_1 + N_2 + N_3)$ .

Критерий качества классификации определялся следующим образом. Для каждого объекта  $x$ , принадлежащего классу  $C_1$  в обучающем множестве данных, вычислялся показатель разделимости классов

$$Q(x) = \mu_1(x) - \max(\mu_2(x), \mu_3(x)),$$

где  $\mu_i(x)$  ( $i=1, 2, 3$ ) суть функции принадлежности  $x$  нечетким классам  $G_i$ :

$$\mu_i(x) = \bigcup_{j=1}^{N_i} \bigcap_{k=1}^5 \mu_{ijk}(x).$$

Чем выше значение  $Q(x)$ , тем лучше классификация объекта  $x$ . Аналогично определялись показатели разделимости для других классов. Для каждого класса  $C_i$  определялось среднее значение разделимости  $Q_i$  по всем его элементам, и сумма этих средних значений определяла общий критерий классификации:  $Q_s = Q_1 + Q_2 + Q_3$ . Максимизация этого критерия по параметрам функций принадлежности применялась для определения искомой нечеткой классификации.

Предложенный метод построения нечетких классов опробован на реальных данных по скважине. В рассматриваемом примере использовалось 326 образцов породы. Эти данные содержали соответственно 59, 128 и 139 образцов из классов  $C_1$ ,  $C_2$  и  $C_3$ . Из каждого класса случайным образом были отобраны 90% образцов в качестве обучающей выборки, оставшиеся 10% вошли в тестирующую выборку. Каждый нечеткий класс  $G_i$  представлялся с помощью трех нечетких гранул  $G_i = G_{i1} \cup G_{i2} \cup G_{i3}$ . Для максимизации критерия классификации применялся алгоритм нелинейной оптимизации с ограничениями из оптимизационного пакета MATLAB. На основе полученной нечеткой классификации результирующая классификация

проводилась следующим образом. Образец относился к классу, степень принадлежности к которому в полученной нечеткой классификации была максимальна. Правильная классификация на обучающих данных составила соответственно 74%, 65% и 88% для классов  $C_1$ ,  $C_2$  и  $C_3$  соответственно. Правильная классификация на тестовых данных составила соответственно 33%, 62% и 86%. Классификация образцов класса  $C_3$  (пористо-трещиноватый) по экспертной оценке была оценена как удовлетворительная, если принять во внимание, что только результаты измерений по пяти параметрам DT4P, RHOV, NPFI, GR и MSFL использовались как основа для классификации пористости пород.

Нечеткие гранулы могут описывать плохо разделяемые на практике классы пористости пород. Полученные нечеткие классы могут использоваться в дальнейшем для построения правил классификации в виде

If ( $p_1(x)$  is  $\mu_{311}(x)$  and ...  $p_5(x)$  is  $\mu_{315}(x)$ ) or  
 ( $p_1(x)$  is  $\mu_{321}(x)$  and ...  $p_5(x)$  is  $\mu_{325}(x)$ ) or  
 ( $p_1(x)$  is  $\mu_{331}(x)$  and ...  $p_5(x)$  is  $\mu_{335}(x)$ )  
 then  $x$  is  $C_3$

Анализ построенных функций принадлежности показал, что, фактически, классификация класса  $C_3$  основана лишь на трех параметрах DT4P, NPFI и GR, так как нечеткие интервалы, определяемые функциями принадлежности гранул по остальным двум параметрам, совпадают с областью определения этих переменных. В результате было построено четкое правило

If  $DT4P(x) > 5.6$  and  $NPFI(x) < 3.1$  and  $GR(x) < 6.2$  then  $x$  is  $C_3$ ,

где указаны нормализованные численные значения параметров.

Правильная классификация по этому правилу составила 95%. Это правило покрывает 78% объектов из класса  $C_3$ .

### Классификация данных с помощью нейронных сетей

Решение задачи классификации пористости пород осуществлялось также с помощью нейронных сетей (НС) прямого распространения [2, 3] с одним и двумя скрытыми слоями (количество нейронов в слое варьировалось от 10 до 30) и вероятностной НС (ВНС) [3]. Результаты классификации приведены в таблице 1. Как показывают результаты проведенных исследований, наилучший результат классификации для рассматриваемой задачи достигнут двухслойной нейронной сетью с 20 нейронами в каждом слое. Архитектура этой нейронной сети представлена на рис. 1, где  $X$  и  $Y$  обозначают вектора входных и выходных переменных системы.

Табл. 1. Результаты классификации для НС различной архитектуры

Количество нейронов в первом слое	Количество нейронов во втором слое	Количество нейронов в выходном слое	Процент ошибки на обучающей выборке	Процент ошибки на тестовой выборке
10	10	1	19,7	17,3
20	20	1	18	15,1
10	-	1	28	17,5
20	-	1	23,3	18,3
30	-	1	18	19

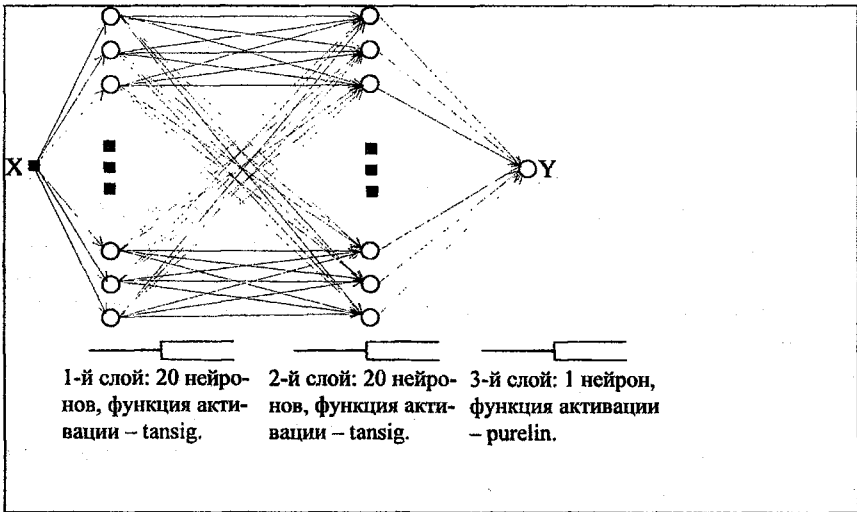


Рис. 1. Архитектура НС прямого распространения с двумя скрытыми слоями

Для классификации данных применялась также вероятностная НС [3], являющаяся сетью прямого распространения (без обратных связей) и состоящая из двух слоев. В качестве функций активации в ВНС применяются не пороговые функции, а функция  $radbas(x) = e^{-2x}$  в первом слое и конкурирующая функция активации, преобразующая вектор входа слоя нейронов так, чтобы нейрон с самым большим входом имел выход, равный 1, а все другие нейроны имели выходы, равные 0, во втором слое. В ВНС заранее не известно количество нейронов в скрытых слоях, оно определяется при обучении НС, когда ей представляется обучающая выборка, а также параметр влияния *spread*. Этот параметр отвечает за диапазон входных значе-

ний, который должен быть принят во внимание каждым из нейронов в первом слое. Его надо выбирать большим, чем шаг разбиения интервала задания обучающей последовательности, но меньшим размера самого интервала. Варьируя величиной параметра, можно уменьшить процент ошибок на тестовой выборке, но тогда увеличивается процент ошибок на обучающей выборке.

Наилучший результат на тестовой выборке был 19,4%, процент ошибок на обучающей выборке при этом составил 9,3% (параметр *spread* был равен 10).

На обучающей выборке можно добиться и нулевого процента ошибок, для этого надо сделать параметр *spread* достаточно малым (примерно равным трем), но тогда процент ошибок на тестовой выборке возрастает до 27%.

### Заключение

Предложенный в работе метод классификации на основе нечетких гранул может использоваться в задачах классификации плохо разделимых классов. В общем случае увеличение в модели числа гранул, образующих нечеткие классы, приводит к улучшению классификации, поскольку позволяет покрывать скопления объектов из некоторого класса, разделенные объектами других классов. Анализ результатов классификации данных с помощью нейронных систем различной архитектуры показал, что наилучший результат достигается при использовании нейронной сети с двумя скрытыми слоями и с 20 нейронами в каждом слое. Сравнение результатов классификации с помощью нечетких гранул и с помощью нейронных сетей показывает, что классификация данных с помощью нейронных сетей дает лучшие результаты, однако содержательная интерпретация архитектуры нейронной сети в терминах используемых признаков объектов затруднительна. Нечеткая классификация с помощью гранул может быть легко представлена в виде правил, что позволяет содержательно интерпретировать классы в терминах признаков.

Работа выполнялась при поддержке проектов IMP CDI.00006 и РФФИ № 03-01-96245.

### Литература

1. Аверкин А.Н., Батыршин И.З., Блишун А.Ф., Силов В.Б., Тарасов В.Б. Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Д.А. Поспелова. - М.: Наука, 1986.
2. Каллан Р. Основные концепции нейронных сетей. - М.: Вильямс, 2001.

3. Медведев В.С., Потемкин В.Г. Нейронные сети. MATLAB6. – М.: ДИАЛОГ-МИФИ, 2002. (Пакеты прикладных программ; Кн. 4).
4. Jang J.-S.R., Sun C.T., Mizutani E. Neuro-Fuzzy and Soft Computing. A Computational Approach to Learning and Machine Intelligence. - Prentice-Hall International, 1997.
5. Kosko B. Fuzzy Engineering. - New Jersey: Prentice-Hall, 1997.
6. Romero-Cortés J., Sheremetov L. Fuzzy Bayesian classifier: a multi-agent system for information retrieval in the Web // Advances in Soft Computing: Neural Networks and Soft Computing (Proceedings of the Sixth International Conference on Neural Network and Soft Computing, Zakopane, Poland, 2002). - Physica-Verlag, 2003. - P. 444-449.
7. Batyrshin I., Sheremetov L., Markov M., Panova A. Logging Data Processing for Pore Structure Classification // Proceedings of The Second Workshop Intelligent Computing in the Petroleum Industry, ICPI-2003. - Acapulco, Mexico, 2003. – P. 41- 46.
8. Kazatchenko E., Markov M., Mousatov A. Determination of primary and secondary porosity in carbonate formations using acoustic data // SPE Annual Conf. 2003. - Denver, Colorado, Paper SPE 84209.