А.В. Авилов, В.Э. Борзых

АЛГОРИТМ КЛАССИФИКАЦИИ ГОРНЫХ ПОРОД НА ЛИТОЛОГО-ПЕТРОФИЗИЧЕСКИЕ ТИПЫ

Статья посвящена разработке алгоритма для решения задачи классификации горных пород на литолого-петрофизические типы на основе технологии нечетких деревьев решений.

Классификация горных пород, литолого-петрофизический тип, нечеткие деревья решений.

Одним из подходов при систематизации литологических и петрофизических параметров является деление пород на типы в зависимости от их структурно-текстурных особенностей, минералогического состава, эпигенетической преобразованности и коллекторских свойств.

Под *литолого-петрофизическим типом пород* (далее — ЛПТП) понимается совокупность пород изучаемого объекта (пласта) с одинаковыми генетическими признаками, близкими литолого-минералогическими и фильтрационно-емкостными характеристиками, которые обусловливают для коллекторов единообразную структуру пустотного пространства (контролируется одинаковой последовательностью процессов минералообразования), для неколлекторов — единую природу потери коллекторских свойств. Полный набор выделенных литолого-петрофизических типов характеризует весь интервал изменения фильтрационно-емкостных свойств пласта [2].

В статье [1] сформулированы принципы и приведена общая схема этапов классификации. Однако представленный в ней подход слишком общий и не учитывает того, что процедура деления на типы пород не является формальной. В пределах каждого объекта классификация проводится с учетом его индивидуальных особенностей, что позволяет находить механизмы, контролирующие фильтрационно-емкостные свойства пород и обосновывать границу «коллектор-неколлектор». Объекты классификации также делятся на группы по времени и условиям образования, видам и интенсивности эпигенетических процессов, протекавших в породах.

Таким образом, имеется задача классификации с входным множеством количественных признаков описаний объектов и нечетким типом классов. В данном случае целесообразно применить технологию нечетких деревьев решений (далее — НДР). Использование этой технологии обусловлено также тем, что НДР:

- позволяют не только сформировать набор классифицирующих правил, но и в дальнейшем применять их для классификации новых образцов;
- определить не только признак принадлежности объекта к одному из классов, но и величину принадлежности (таким образом не теряется информация о том, что объект может обладать свойствами разных признаков в той или иной мере) [3];
- являются легко интерпретируемыми и понятными, так как результаты работы представляются в виде иерархической последовательной структуры [3];
- обладают высокой точностью классификации, достигаемой за счет сочетания достоинств нечеткой логики и деревьев решений [3].

При оперировании выборками входных данных большого объема, единожды обучить и сформировать дерево, обеспечивающее приемлемый процент корректной классификации, не представляется возможным. Решением данной проблемы является организация расширения дерева решений. Для этого при каждой классификации нового объекта необходимо заменять имеющееся дерево решений на новое при условии обеспечения им большей эффективности классификации, а значит, необходимо сохранять не только классифицирующие правила, но и обучающие выборки. Алгоритм классификации нового объекта в общем виде можно представить следующим образом (рис. 1).

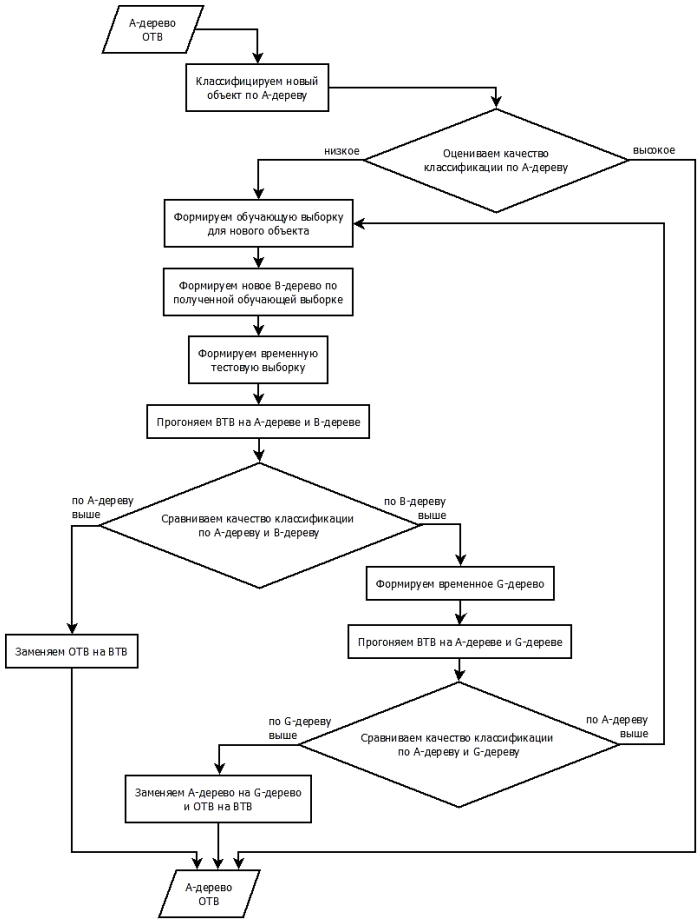

Рис. 1. Схема алгоритма классификации нового объекта

1. Пусть имеется главное дерево классификации (далее — α-дерево) и общая тестовая выборка данных с проставленными ЛПТП (далее — ОТВ).
2. Классифицируем новый объект по α-дереву — для каждого образца обходим дерево сверху вниз до выявления соответствующего ЛПТП.
3. Оцениваем качество классификации. Если качество классификации неприемлемое, то переходим к шагу 3.
4. Формируем обучающую выборку для нового объекта, проставляем ЛПТП вручную либо автоматически на основе имеющихся близких значений ОТВ.

5. Формируем новое дерево по полученной обучающей выборке (далее — β-дерево).

6. Формируем временную тестовую выборку (далее — ВТВ) путем расширения ОТВ полученной обучающей выборкой.

7. Прогоняем ВТВ на α- и β-дереве.

8. Сравниваем α- и β-дерево. Если качество классификации по α-дереву выше такового по β-дереву, то оставляем α-дерево без изменений и заменяем ОТВ на ВТВ, иначе переходим к шагу 9.

9. Формируем новое дерево по ВТВ (далее — γ-дерево), используя в качестве выходной переменной ЛПТП α-дерева в случае его выбора на предыдущем шаге либо ЛПТП β-дерева в случае его выбора на предыдущем шаге.

10. Прогоняем ВТВ на α- и γ-дереве.

11. Сравниваем γ- и α-дерево. Если качество классификации по γ-дереву выше такового по α-дереву, то заменяем последнее на γ-дерево и ОТВ на ВТВ, иначе делаем вывод о некорректности обучающей выборки нового объекта и переходим к шагу 4.

Рассмотрим алгоритм на тестовом примере пластовой группы АС.

1. Пусть имеется ОТВ, состоящая из 15 образцов (табл. 1), и α-дерево с классифицирующими правилами (рис. 2).

Таблица 1

Образец	Кп	Кпр	Квс	Роб	Рмин	ЛПТП
1000-00	21,052	282,440	25,088	2,087	2,643	1
1001-00	19,360	17,350	42,090	2,150	2,660	1
1002-00	22,970	129,150	31,114	2,054	2,666	1
1003-00	22,659	45,970	34,707	2,052	2,653	3
1004-00	20,982	13,810	44,196	2,103	2,661	3
1005-00	20,135	1,430	67,720	2,140	2,679	3
1006-00	17,671	0,730	72,939	2,202	2,675	4
1007-00	17,600	0,130	88,196	2,231	2,707	4
1008-00	17,374	0,120	86,136	2,213	2,678	4
1009-00	11,708	0,590	55,506	2,351	2,662	5
1010-00	11,878	0,620	55,190	2,345	2,661	5
1011-00	13,770	0,520	66,928	2,300	2,670	5
1012-00	19,689	5,390	49,706	2,145	2,673	2
1013-00	18,400	3,690	59,700	2,180	2,680	2
1014-00	18,800	4,540	59,900	2,170	2,680	2

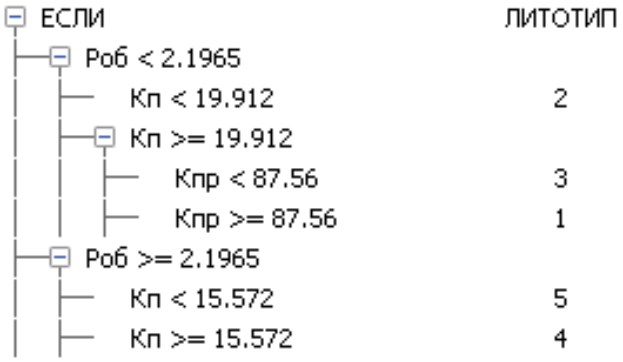

Рис. 2. α-дерево

2. Классифицируем новый объект, состоящий из 10 образцов (табл. 2), по имеющемуся α-дереву.

Таблица 2

Образец	Кп	Кпр	Квс	Роб	Рмин	ЛПТП по α-дереву
2000-01	21,509	139,670	28,668	2,078	2,647	1
2001-01	21,760	6,110	47,460	2,090	2,670	3
2002-01	19,140	1,360	53,600	2,160	2,660	2
2003-01	14,820	0,020	93,840	2,290	2,690	5
2004-01	15,940	16,500	38,850	2,240	2,670	4
2005-01	15,570	3,280	52,130	2,250	2,660	5
2006-01	13,720	0,060	92,400	2,330	2,700	5
2007-01	14,850	0,200	86,040	2,280	2,670	5
2008-01	5,920	0,090	83,170	2,530	2,680	5
2009-01	12,200	0,470	59,670	2,340	2,660	5

3. Оценим качество классификации по α-дереву (табл. 3). По α-дереву с низкой достоверностью классифицировано 7 из 10 образцов, необходимо перейти к шагу 4.

Таблица 3

№	Условие			Следствие	Поддержка		Достоверность	
	Показатель	Знак	Значение	ЛПТП	Кол-во	%	Кол-во	%
1	Кп	<	17,54	5	7	70,00	6	85,71
2	Кп	>=	17,54	1	3	30,00	1	33,33

4. Пусть имеется обучающая выборка для нового объекта (табл. 4).

Таблица 4

Образец	Кп	Кпр	Квс	Роб	Рмин	ЛПТП
2000-01	21,509	139,670	28,668	2,078	2,647	1
2001-01	21,760	6,110	47,460	2,090	2,670	1
2002-01	19,140	1,360	53,600	2,160	2,660	2
2003-01	14,820	0,020	93,840	2,290	2,690	2
2004-01	15,940	16,500	38,850	2,240	2,670	3
2005-01	15,570	3,280	52,130	2,250	2,660	3
2006-01	13,720	0,060	92,400	2,330	2,700	4
2007-01	14,850	0,200	86,040	2,280	2,670	4
2008-01	5,920	0,090	83,170	2,530	2,680	5
2009-01	12,200	0,470	59,670	2,340	2,660	5

5. Формируем β-дерево (рис. 3).

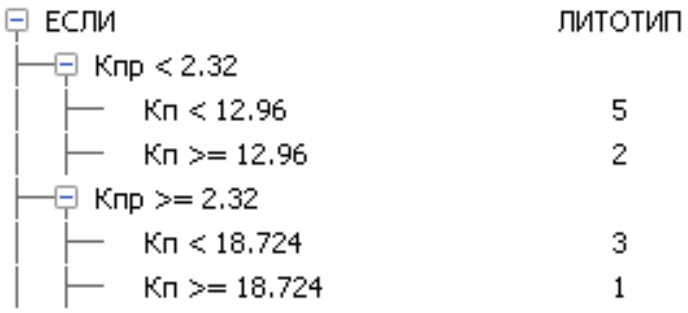

Рис. 3. β-дерево

6. Формируем ВТВ (табл. 5).

Таблица 5

Образец	Кп	Кпр	Квс	Роб	Рмин	ЛПТП	ЛПТП по α-дереву	ЛПТП по β-дереву
1000-00	21,052	282,440	25,088	2,087	2,643	1	1	1
1001-00	19,360	17,350	42,090	2,150	2,660	1	1	1
1002-00	22,970	129,150	31,114	2,054	2,666	1	1	1
1003-00	22,659	45,970	34,707	2,052	2,653	3	3	1
1004-00	20,982	13,810	44,196	2,103	2,661	3	3	1
1005-00	20,135	1,430	67,720	2,140	2,679	3	3	2
1006-00	17,671	0,730	72,939	2,202	2,675	4	4	2
1007-00	17,600	0,130	88,196	2,231	2,707	4	4	2
1008-00	17,374	0,120	86,136	2,213	2,678	4	4	2
1009-00	11,708	0,590	55,506	2,351	2,662	5	5	5
1010-00	11,878	0,620	55,190	2,345	2,661	5	5	5
1011-00	13,770	0,520	66,928	2,300	2,670	5	5	2
1012-00	19,689	5,390	49,706	2,145	2,673	2	2	1
1013-00	18,400	3,690	59,700	2,180	2,680	2	2	3
1014-00	18,800	4,540	59,900	2,170	2,680	2	2	1
2000-01	21,509	139,670	28,668	2,078	2,647	1	1	1
2001-01	21,760	6,110	47,460	2,090	2,670	1	3	1
2002-01	19,140	1,360	53,600	2,160	2,660	2	2	2
2003-01	14,820	0,020	93,840	2,290	2,690	2	5	2
2004-01	15,940	16,500	38,850	2,240	2,670	3	4	3
2005-01	15,570	3,280	52,130	2,250	2,660	3	5	3
2006-01	13,720	0,060	92,400	2,330	2,700	4	5	2
2007-01	14,850	0,200	86,040	2,280	2,670	4	5	2
2008-01	5,920	0,090	83,170	2,530	2,680	5	5	5
2009-01	12,200	0,470	59,670	2,340	2,660	5	5	5

7. Прогоняем ВТВ на α- и β-дереве (предпоследний и последний столбец табл. 5).

8. Сравниваем качество классификации по α-дереву (табл. 6) и β-дереву (табл. 7). По α-дереву с высокой достоверностью классифицировано 23 из 25 образцов, по β-дереву — со 100 % достоверностью все образцы. Таким образом, качество классификации по β-дереву выше качества классификации по α-дереву, необходимо перейти к шагу 9.

Таблица 6

№	Условие			Следствие	Поддержка		Достоверность	
	Показатель	Знак	Значение	ЛПТП	Кол-во	%	Кол-во	%
1	Роб	<	2,245	4	4	16,67	4	100,00
	Кп	<	18,035					
2	Роб	<	2,245	2	4	16,67	4	100,00
	Кп	>=	18,035					
	Кпр	<	5,75					
3	Роб	<	2,245	3	2	8,33	2	100,00
	Кп	>=	18,035					
	Кпр	>=	5,75					
	Кпр	<	15,58					
4	Роб	<	2,245	1	5	20,83	4	80,00
	Кп	>=	18,035					
	Кпр	>=	15,58					
5	Роб	>=	2,245	5	9	37,50	9	100,00

Таблица 7

№	Условие			Следствие	Поддержка		Достоверность	
	Показатель	Знак	Значение	ЛПТП	Кол-во	%	Кол-во	%
1	Кп	<	12,96	5	4	16,67	4	100,00
2	Кп	>=	12,96	2	9	33,33	9	100,00
	Кпр	<	2,32					
3	Кп	>=	12,96	3	3	12,50	3	100,00
	Кпр	>=	2,32					
	Кп	<	18,6					
4	Кпр	>=	2,32	1	9	37,50	9	100,00
	Кп	>=	18,6					

9. Формируем γ-дерево (рис. 4).

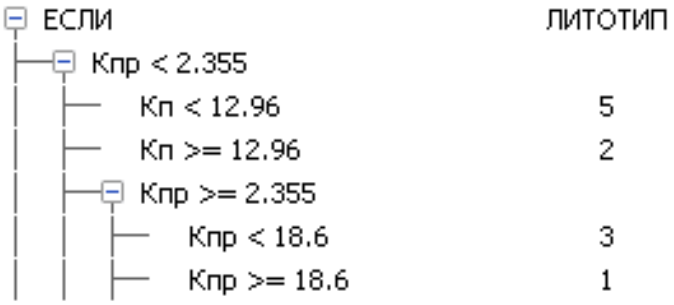

Рис. 4. γ-дерево

10. Прогоняем ВТВ на α- и γ-дереве (табл. 8).

11. Сравниваем качество классификации по α-дереву (табл. 6) и γ-дереву (табл. 9). По γ-дереву со 100 % достоверностью классифицированы все образцы. Таким образом, качество классификации по γ-дереву выше качества классификации по α-дереву, заменяем α-дерево на γ-дерево и ОТВ на ВТВ.

Таблица 8

Образец	Кп	Кпр	Квс	Роб	Рмин	ЛПТП	ЛПТП по α-дереву	ЛПТП по γ-дереву
1000-00	21,052	282,440	25,088	2,087	2,643	1	1	1
1001-00	19,360	17,350	42,090	2,150	2,660	1	1	1
1002-00	22,970	129,150	31,114	2,054	2,666	1	1	1
1003-00	22,659	45,970	34,707	2,052	2,653	3	3	1
1004-00	20,982	13,810	44,196	2,103	2,661	3	3	1
1005-00	20,135	1,430	67,720	2,140	2,679	3	3	2
1006-00	17,671	0,730	72,939	2,202	2,675	4	4	2
1007-00	17,600	0,130	88,196	2,231	2,707	4	4	2
1008-00	17,374	0,120	86,136	2,213	2,678	4	4	2
1009-00	11,708	0,590	55,506	2,351	2,662	5	5	5
1010-00	11,878	0,620	55,190	2,345	2,661	5	5	5
1011-00	13,770	0,520	66,928	2,300	2,670	5	5	2
1012-00	19,689	5,390	49,706	2,145	2,673	2	2	1
1013-00	18,400	3,690	59,700	2,180	2,680	2	2	3
1014-00	18,800	4,540	59,900	2,170	2,680	2	2	1
2000-01	21,509	139,670	28,668	2,078	2,647	1	1	1
2001-01	21,760	6,110	47,460	2,090	2,670	1	3	1
2002-01	19,140	1,360	53,600	2,160	2,660	2	2	2

Окончание т а б л . 8

Образец	Кп	Кпр	Квс	Роб	Рмин	ЛПТП	ЛПТП по α-дереву	ЛПТП по γ-дереву
2003-01	14,820	0,020	93,840	2,290	2,690	2	5	2
2004-01	15,940	16,500	38,850	2,240	2,670	3	4	3
2005-01	15,570	3,280	52,130	2,250	2,660	3	5	3
2006-01	13,720	0,060	92,400	2,330	2,700	4	5	2
2007-01	14,850	0,200	86,040	2,280	2,670	4	5	2
2008-01	5,920	0,090	83,170	2,530	2,680	5	5	5
2009-01	12,200	0,470	59,670	2,340	2,660	5	5	5

Т а б л и ц а 9

№	Условие			Следствие	Поддержка		Достоверность	
	Показатель	Знак	Значение	ЛПТП	Кол-во	%	Кол-во	%
1	Кпр	<	2,355	5	3	12,50	3	100,00
	Кп	<	12,96					
2	Кпр	<	2,355	2	9	37,50	9	100,00
	Кп	>=	12,96					
3	Кпр	>=	2,355	3	4	12,50	4	100,00
	Кп	<	18,6					
4	Кпр	>=	2,355	1	9	37,50	9	100,00
	Кп	>=	18,6					

Применение данного алгоритма обеспечивает:

• комплексирование широкого спектра методов исследований петрофизических и литологических свойств горных пород;

• сокращение времени классификации;

• повышение точности классификации за счет сочетания достоинств нечеткой логики и деревьев решений;

• независимость качества классификации от квалификации сотрудника.

ЛИТЕРАТУРА

1. *Авилов А.В., Борзых В.Э., Федорцов И.В.* Автоматизация процесса анализа результатов исследований керна и классификации пород // Вестн. кибернетики. Тюмень: Изд-во ИПОС СО РАН, 2011. № 10. [Электрон. ресурс]. Режим доступа: http://www.ipdn.ru.

2. *Коровина Т.А., Кропотова Е.П.* Форма представления литологической информации при интерпретации ГИС и пластовом моделировании // Пути реализации нефтегазового потенциала ХМАО: Материалы III Науч.-практ. конф. Ханты-Мансийск, 2000. С. 346–351.

3. *Царьков С.* Нечеткие деревья решений. Технологии анализа данных BaseGroup Labs [Электрон. ресурс]. Режим доступа: http://www.basegroup.ru/library/analysis/fuzzylogic/fuzzy_dtrees/.

A.V. Avilov, V.E. Borzykh

ALGORITHM OF CLASSIFYING GEOLOGICAL MATERIAL INTO LITHOLOGIC AND PETROPHYSICAL TYPES

The article is devoted to the development of algorithm to solve a problem of classification of geological material into lithologic and petrophysical types, basing on a technology of fuzzy decision trees.

Classification of geological material, lithologic and petrophysical type, fuzzy decision trees.