

基于粗糙集和 Petri 网的油层含油识别研究

张漫¹, 李晶莹², 严胡勇³, 王梅¹, 范广玲⁴

- (1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318;
2. 中国石油化工股份有限公司江苏油田采油一厂, 江苏 扬州 225200;
3. 重庆工商大学 计算机科学与信息工程学院, 重庆 400067;
4. 东北石油大学 数学与统计学院, 黑龙江 大庆 163318)

摘要:油藏受天然气的侵入,就会产生油气分异的现象;天然气会萃取出油藏中的轻组分,造成凝析油藏或轻质油藏;此外,油藏中的重组分脱出,形成沉淀,直接导致储层的渗透率降低,从而影响油气的分布规律。针对油藏中油气分布规律复杂的问题,在传统油层含油判别分析的基础上,提出了一种基于粗糙集和 Petri 网的油藏建模方法。应用粗糙集的知识约简对油层含油识别相关的岩层厚度、泥质含量等 6 个指标进行属性选择,提取最简规则,建立 Petri 网模型,根据 Petri 网的并行推理达到简洁高效的含油识别。仿真实验结果表明,采用粗糙集与 Petri 网判断的油气分布规律与现场数据高度接近,精度高,识别速度快,正确率高。可见将粗糙集与 Petri 网组合用于油层油气识别是有效的。

关键词:粗糙集;Petri 网;油气分布;识别

中图分类号:TE19

文献标识码:A

文章编号:1673-629X(2020)01-0174-05

doi:10.3969/j.issn.1673-629X.2020.01.031

Recognition at Oil and Gas Distribution in Oilfield Based on Rough Sets and Petri Nets

ZHANG Man¹, LI Jing-ying², YAN Hu-yong³, WANG Mei¹, FAN Guang-ling⁴

- (1. School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China;
2. No. 1 Oil Production Plant, Jiangsu Oil Field of China Petroleum Chemical Corporation, Yangzhou 225200, China;
3. School of Computer Science and Information Engineering, Chongqing Technology and Business University, Chongqing 400067, China;
4. School of Mathematics and Statistics, Northeast Petroleum University, Daqing 163318, China)

Abstract: If the reservoir is invaded by foreign natural gas, it will produce oil and gas differentiation; on the one hand, the light component of the reservoir will be extracted by natural gas to result in condensate reservoir or light oil reservoir; on the other hand, the recombination of the reservoir will be separated to result in solid precipitation. These precipitates will reduce the permeability of the reservoir and even affect the distribution of oil and gas. Aiming at the complex distribution law of oil and gas in oil deposit, based on the discriminant analysis of oiliness in traditional oil layer, we put forward a modeling approach for oil deposit on the basis of rough set and Petri net. By applying the simple knowledge in rough set, attribute optimization could be realized towards the thickness of stratum, shale content and other 4 indexes relevant to the oiliness identification of oil layer, and the minimum identification rule could be acquired for establishing Petri net model. Through parallel inference of Petri net, the concise and high-efficient oiliness identification could be realized. The simulation shows that the distribution law of oil and gas judged after adopting rough set and Petri net is highly close to the practical situation, with fast the identification speed high accuracy. It is obvious that using the combination of rough set and Petri net to identify the oil and gas in oil layer is effective.

Key words: rough sets; Petri nets; oil and gas distribution; recognition

0 引言

油藏的形成是动态的地质条件与时间、空间相匹配的结果,它是一个系统的过程。随着石油勘探领域的不断发展,含油层识别的研究对象越来越多,加之地质条件的复杂性,使得油层的含油识别难度也相应增大。国内的学者^[1-2]对含油气性进行识别时所考虑的侧井属性也不尽相同。冯国庆等^[3]选取参数孔隙度(POR)、声波时差(AC)、含水饱和度(Sw)、感应测井值(COND)4个指标来识别储层油性;傅强等^[4]根据测井处理的特点,选择易于获取的含水饱和度(Sw)、声波(AC)、泥质含量(Vsh)、渗透率(PERM)、孔隙度(POR)、电阻率(RT)6种参数进行储层识别;裴亦楠等^[5]将冲洗带含水饱和度(SWM)、自然电位减少系数(ALF)、地层径向电阻率比值(RTXO)、视地层水电阻率(RWA)、地层电阻率增大率(RTO)和地层电阻率相对值(RTI)用于评价储层含油性的指标。目前的含油气识别研究主要是通过学者自行选定的、认为合理的属性作为输入,对主观经验依赖比较大,较少地给出客观的原则或依据,通过主观选定认为合理的输入变量,指标的选择存在较大的经验性。对含油性识别相关的测井属性的选择不仅是由少到多的过程,也是由多到少的过程。这里的由少到多的过程指尽可能提取到与含气性相关度比较大的测井属性;而由多到少的过程是指结合具体的研究目的,从测井属性中筛选出最佳的属性子集。文中针对油藏中油气分布规律复杂的问题,在传统油层含油判别分析的基础上,提出了一种基于粗糙集与 Petri 网的油藏建模方法。应用粗糙集的知识约简对油层含油识别相关的岩层厚度、泥质含量等指标进行属性选择,提取最简规则,建立 Petri 网模型,根据 Petri 网的并行推理达到简洁高效的含油识别。

粗糙集理论(rough sets theory)是由波兰科学家 Z. Pawlak 教授于 1982 年提出的一种处理不精确、不一致、不完整等不完备信息的数学工具。粗糙集有两个特点:一是无需先验知识,二是易用性强。因此,粗糙集已广泛应用于模式识别与分类、图像处理、临床医疗诊断、工业过程故障诊断等多个领域^[6-10]。由于粗糙集理论创建的目的以及研究的出发点在于直接对原始数据进行分析与推理,从而发现隐藏的知识,揭示潜在的规律,因此粗糙集理论是一种天然的数据挖掘或知识发现方法。与传统的基于证据理论的数据挖掘方法、基于模糊理论的数据挖掘方法和基于概率论的数据挖掘方法等其他处理不确定性问题理论的方法相比,最大的区别在于粗糙集理论不要求解问题所需的先验知识,此外,与其他处理不确定性问题的理论、方法有着较强的互补性。粗糙集能够在保留关键信息

的前提下对知识进行处理,并求得知识的最小表达^[11]。它可分析不完整数据,找出数据间的关系,从而提取有用的属性,此方法克服了传统方法评价参数选择的主观性,但也存在着并行推理能力不足的缺点。而 Petri 网是对离散系统的并行数学表达,特别适用于描述异步、并发的系统模型。它具有强大的并行推理能力,但存在不能选择条件属性等缺点。因此,将粗糙集与 Petri 网相结合解决油层含油识别问题具有较强的互补性,可以得到较好的识别效果。

1 基本理论

1.1 粗糙集理论

定义 1: 设一个四元组 $IS = (U, A, V, f)$ 为决策表,其中 U 为论域, A 为属性集,包括条件属性 C 与决策属性 D , 且 $C \cap D = \emptyset$, $C \cup D = A$, 具有条件属性与决策属性的知识表达系统称为决策表, V 为属性值域集, $f: U \times A \rightarrow V$ 代表决策表的一个信息函数,每个信息对象的每个属性都被赋予了一个值。

定义 2: 在知识表达系统 $IS = (U, A, V, f)$ 中,每个子集 $X \subseteq U$ 和一个等价关系 $R \subseteq A$, 那么子集 $RX = \{x \in U \mid [x]_R \subseteq X\}$ 是 X 的 R 下近似集,即正域,记为 $POS_R(X)$ 。

定义 3: $IS = (U, A, V, f)$ 为一个知识表达系统, $P, Q \in A$, 称子集 $POS_P(Q) = \bigcup_{X \in U/Q} PX$ 为 Q 的 P 正域。

定义 4: 设 $IS = (U, A, V, f)$ 为一个知识表达系统, $P, Q \in A, R \in P$, 如果 $POS_P(Q) = POS_{(P - \{R\})}(Q)$, 那么 R 在 P 中相对于 Q 不必要,否则就是必要的。粗糙集中最核心的问题就是属性约简,属性约简就是在保持决策表分类能力不变的前提下,通过对知识的简化求解出问题的分类或决策规则^[12]。

定义 5: 设 $IS = (U, A, V, f)$ 为一个知识表达系统, $P, Q \in A$, 若每个 $R \in P$ 在 P 中相对于 Q 都是必要的,那么称 P 为 Q 独立,否则,称为依赖的。对相依赖的属性,其中包含有多余属性,可对其约简。

1.2 Petri 网原理

Petri 网是联邦德国的卡尔·亚当·帕奇(Carl Adam. Petri)于 1962 年在他的博士论文《用自动机通讯》中提出的数学模型。Petri 网是一种异步系统并行建模与分析的重要工具,它主要关注于系统中各事件间发生变化的原因、经过以及发生后彼此间的相互变动等^[13]。由于 Petri 网是网状结构,这样就使它能在描述异步并行行为上有着巨大的优势,结合其优异的图形表达能力,使得它能在多个领域中得到广泛的应用。Petri 网并非描述自然界中已经存在的自然规律,它所描述的是一个具有反应联系的模型,依靠网状结构的特性, Petri 网能够较好地分析出规律间的依赖关

系,从而能客观反映事物间的相互推进关系,使得它具有有良好的易于实现性。

通过 Petri 网模型来描述离散事件系统,可以形成一个简单的 Petri 网^[14]。Petri 网是由变迁 T (方框)、库所 S (圆)、连接库所与变迁的有向弧以及初始托肯 (Token) 组成的一个有向网络。库所主要用来容纳事件中的有效资源。变迁体现的是事件中一个状态到另一状态的动态变化,通常是不可变化的。有向弧是库所到变迁的联系状态,有单向的也有双向的,有时还会注明有向弧的权值(一次性消耗资源数)。托肯表示的是系统资源,是对系统中所在库所的一种动态描述。库所与托肯可以理解为:如果库所中有一个托肯,那么表示此库所有且只能实现一次;如果库所中没有托肯,那么该库所就不能实现。

Petri 网模型原理可以这样描述:输入库所如果有托肯的话,那么意味着从一开始就会被点火,其托肯数量也会相应减少,并通过有向弧向下一个库所传递,由于托肯不断地被传递,托肯初始的位置也会相应地动态变化,进而使得模型能够得到运行。从上述过程可以看出,输入库所是变迁的起点,也是模型运行的前提,而输出库所是变迁的终点,也是模型最后所得到的结论。Petri 网由输入库所通过变迁再到输出库所所经过的完整通路,可表示为一个事件从发生到结束的规则。

定义一个六元组 $\Sigma = (S, T, F, M_0, M_1, U, C)$ 为一个有限 S/T 系统,其中库所节点集合 $S = \{s_1, s_2, \dots, s_m\} (m \geq 0)$, 变迁集合 $T = \{t_1, t_2, \dots, t_n\} (n \geq 0)$, 满足: $S \cap T = \emptyset, S \cup T \neq \emptyset$, 流关系 $F \subseteq (S \times T) \cup (T \times S)$, $\text{dom}(F) \cup \text{cod}(F) = S \cup T$ 。 M_0 是初始托肯, M_1 是变迁被点火后的状态,库所与变迁之间的关系用关联矩阵 C 表示为:

$$C_{ij} = w(t_j - p_i) - w(p_i - t_j)$$

其中, $1 \leq i \leq n, 1 \leq j \leq m$, w 是权函数。

$U = \{u_1, u_2, \dots, u_n\}$ 为转移控制矢量,代表系统变迁点火序列,此时的系统状态方程为:

$$M_1 = M_0 + C \times U$$

托肯的变化可以解释 Petri 网的动态行为,而托肯的变化需要变迁的点火即触发来实现。变迁变化的规律为:(1)若变迁的每个输入库所都有一个托肯,那么此变迁被点火;(2)变迁的点火过程是从它的每个输入库所取出托肯,同时,将该托肯放到它的输出库所中;(3)若变迁中有竞争关系,那么,由环境信息决定托肯的最终归属^[15]。

2 油层识别模型研究

粗糙集-Petri 网模型如图 1 所示。该模型主要分

为两部分:一部分是通过训练样本得到模型参数,另一部分是将测试样本代入模型得到结果。首先将输入数据划分为训练样本与检测样本两部分,再把训练样本的特征属性进行离散化预处理,然后通过约简算法对预处理后的离散数据进行属性约简,在保持分类能力不变的前提下,消除冗余属性,从而减少数据挖掘的工作量,再通过 Petri 推理对约简后的样本进行学习。最后,再用训练好的 Petri 网对检测样本进行分类,就可以实现目标识别。

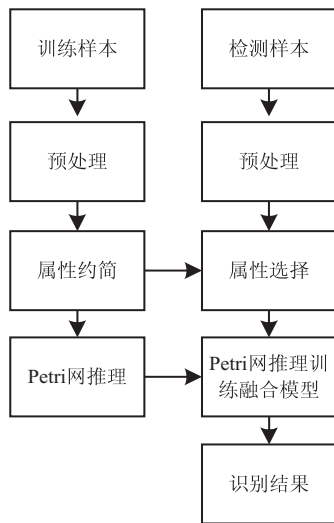


图 1 粗糙集-Petri 网模型

(1)模型参数的选择。油田录井是通过油藏中油层的孔隙度、厚度等参数对油气井水分布规律做出分析的过程,因此,根据油井录井资料中的数据组合解释油气水层是一个模式识别问题^[16],而油气水层的识别是利用多种对其敏感的测井参数进行判别的过程。

(2)数据预处理。在许多实际应用中,粗糙集不能直接从原始的信息系统获取知识,因此,必须对信息系统进行预处理,其中数据离散化是预处理中关键的一步。对信息系统的有效离散化能够大幅提高后续学习过程的效率和学习结果的性能。离散化方法比较多,大致有以下几类:等距和等频法^[17]、布尔量推理离散法^[18]、自适应离散算法^[19]、基于熵的离散化算法^[20]、基于统计检验的离散化算法^[21]、单规则离散器^[22]等等。

(3)建立油层识别系统。以油田某区的油藏样本作为 U ,油气类别作为决策属性,影响油气类别的因素作为条件属性,建立带决策的油层信息识别系统。

(4)属性约简。根据粗糙集理论对各评价属性的重要性进行判定,得到重要属性的最少组合,降低数据维度。

(5)Petri 网知识推理。虽然粗糙集与 Petri 网都能够建立油层含油识别规则,通过粗糙集可以剔除掉不重要的属性,简化判别规则,提高效率;然而,当粗糙

集进行推理时,如果数据量很大,对知识的查表搜索是一个比较费时的过程。若结合 Petri 网进行知识的并行推理,其空间搜索率会得到明显提高,并且是采用图形来表达知识,使推理更加简单清晰。

3 应用实例

文中选取了大庆油田某地区的油藏资料共计 20 个样本进行油气水分布规律实证研究。

3.1 初始评价参数选择

由于该地区油层分布较为复杂,层位也不稳定,油水界面也不统一,并且井间连接性差,基本不连通。因此,就不能简单地确定哪层是否利于开采,哪层不利于开采。而应对油层进行分小区认识。将此地区 20 口油井不同深度中的砂岩厚度 (c_1)、泥质含量 (c_2)、有

效孔隙度 (c_3)、含水饱和度 (c_4)、含油气饱和度 (c_5)、含油气孔隙度 (c_6) 6 个参数作为初始评价参数 (见表 1),把这 20 个样本集合用 $U = \{u_1, u_2, \dots, u_{30}\}$ 表示,这 6 个参数作为样本集合 U 上的条件属性 $C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$,然后通过粗糙集理论分析这 6 个条件属性对解释结果的影响。

3.2 数据预处理

由于粗糙集只能处理离散数据,而这 20 个样本数据都是连续数据,因此需要对这些连续数据进行离散化处理。文中选用等频离散法,将所有条件属性根据频度分为低、中、高三个等级,离散区间见表 2。对于决策属性,将 4 种不同的油气类别量化为 1、2、3、4。类别 1 为水层,类别 2 为干层,类别 3 为差油层,类别 4 为油层。

表 1 初始评价原始数据

样本序号	层厚/m	泥质含量/%	有效孔隙度/%	含水饱和度和/%	含油气饱和度和/%	含油气孔隙度和/%	解释结果
1	0.8	13.66	16.69	71	29	5.63	1
2	0.8	13.15	15.67	68.83	31.17	5.23	1
3	1.8	5.48	11.66	59.69	40.31	5.18	3
4	2.2	11.21	12.71	71.59	28.41	3.55	1
5	1	17.38	12.44	90.32	9.68	1.46	1
...

表 2 指标离散区间

分级	层厚/m	泥质含量/%	有效孔隙度/%	含水饱和度和/%	含油气饱和度和/%	含油气孔隙度和/%
低(1)	(0,1]	(0,9.99]	(0,10.66]	(0,38.13]	(0,31.17]	(0,4.85]
中(2)	(1,1.8]	(9.99,14.85]	(10.66,13.74]	(38.13,64.61]	(31.17,57.23]	(4.85,7.37]
高(3)	>1.8	>14.85	>13.74	>64.61	>57.23	>7.37

3.3 属性约简

运用粗糙集 Semi-minimal 约简算法通过 rosetta 软件实现,得到决策表的一个约简属性集合: {泥质含量,含水饱和度}。这两种指标对油气类别有重要影响,而层厚、有效孔隙度、含油气饱和度与含油气孔隙度对油气类别影响不大,是冗余属性,在下一步分析中可以去除这四项指标。

利用粗糙集属性约简算法对油层识别的分析表明,粗糙集具有强大的知识发现能力,能够发现影响油气类别的主要因素并提取出来,去除冗余条件属性,从而简化数据空间的维数,有利于提高识别的运算速度与精度。

3.4 建立 Petri 网推理模型

由约简后的决策表建立基于 Petri 网的含油识别知识推理模型,如图 2 所示。其中,输入库所 $s_1 - s_6$ 分别表示决策表的最小知识规则,表现为条件形式,如果条件为假,则没有托肯,反之,则库所得到托肯。该模型命题如下: $s_1:c_1 = 1$; $s_2:c_1 = 2$; $s_3:c_1 = 3$; $s_4:c_2 = 1$;

$s_5:c_2 = 2$; $s_6:c_2 = 3$ 。输出库所 $s_7:D = 1$; $s_8:D = 2$; $s_9:D = 3$; $s_{10}:D = 4$ 。

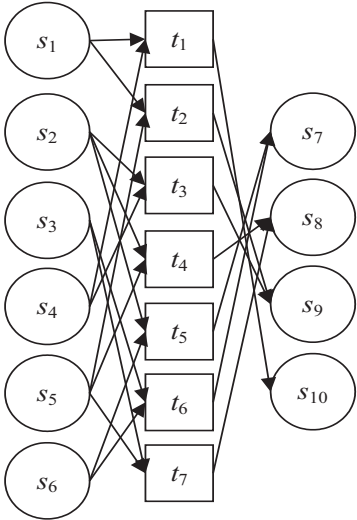


图 2 Petri 网油层含油识别模型

从图 2 可见,经粗糙集约简后,所建立的 Petri 网的知识识别规则删除了冗余信息,推理、识别更为

迅速。

3.5 样本测试

该方法经多个测试样本验证测试,其识别能力较高,并且具有一定的容错能力,因此,该方法是可行的。以下是一些具体样本的测试结果:

样本一:某样本层厚 1 米,泥质含量 7.34%,有效孔隙度 18.5%,含水饱和度 37.34%,含油气饱和度 62.66%,含油气孔隙度 11.61%,其对应约简后的属性值分别为 1,1。将此优化信息输入 Petri 网识别模型,变迁点火后得到终态托肯。测试结论:样本为油层,识别正确。

样本二:某样本层厚 2.6 米,泥质含量 20.27%,有效孔隙度 17.69%,含水饱和度 93.23%,含油气饱和度 6.77%,含油气孔隙度 1.3%,其对应约简后的属性值分别为 3,3。将此优化信息输入 Petri 网识别模型,变迁点火后得到终态托肯。测试结论:样本为水层,识别正确。

样本三:某样本层厚 0.8 米,泥质含量 13.87%,有效孔隙度 9.45%,含水饱和度 49.83%,含油气饱和度 50.17%,含油气孔隙度 5.47%,其对应约简后的属性值分别为 2,2。将此优化信息输入 Petri 网识别模型,变迁点火后得到终态托肯。测试结论:样本为干层,识别正确。

样本四:某样本层厚 1.7 米,泥质含量 6.80%,有效孔隙度 4.42%,含水饱和度 45.11%,含油气饱和度 54.26%,含油气孔隙度 6.18%,其对应约简后的属性值分别为 2,1。将此优化信息输入 Petri 网识别模型,变迁点火后得到终态托肯。测试结论:样本为差油层,识别正确。

4 结束语

文中采用 Petri 网和粗糙集相结合的油层含油识别知识推理模型并将其用于实际,基本概念清楚,逻辑思维严密。通过粗糙集得到知识系统的最简表达决策表,建立了基于知识规则的 Petri 网含油识别推理模型,大幅度减少了规则知识库中的条件维数,有效降低了规则库的规模。测试结果表明,该方法是可行的,并且对复杂样本也具有较好的识别能力,识别能力与速度都比较理想。

参考文献:

- [1] 赵军龙,李甘,朱广社,等.低阻油层成因机理及测井评价方法综述[J].地球物理学进展,2011,26(4):1334-1342.
- [2] 李国欣,欧阳健,周灿灿,等.中国石油低阻油层岩石物理研究与测井识别评价技术进展[J].中国石油勘探,2006,

- 11(2):43-50.
- [3] 冯国庆,李允,谈德辉.模糊贴近度在储层识别中的应用[J].西南石油大学学报,1999,21(4):46-49.
- [4] 傅强,王家林.自组织特征映射网络在储层识别中的应用[J].同济大学学报:自然科学版,1999,27(3):371-374.
- [5] 裴亦楠.油气储层评价技术[M].西安:石油工业出版社,1997.
- [6] 张漫,严胡勇,翟兴然.粗糙集在优化油田开发效果综合评价指标中的应用[J].黑龙江八一农垦大学学报,2017,29(3):96-100.
- [7] 严胡勇,王国胤,张学睿,等. A fast method to evaluate water eutrophication [J]. 中南大学学报:英文版,2016,23(12):3204-3216.
- [8] 严胡勇,王国胤,张学睿,等.粗糙集理论在水质富营养化分级指标权重确定中的应用[C]//第二届全国流域生态保护与水污染控制研讨会论文集.北京:中国环境科学学会,2014:1-10.
- [9] 封丽,张学睿,封雷,等.基于粗糙集的三峡库区支流水质富营养化模糊综合评价模型研究[J].环境工程,2015,33(12):105-110.
- [10] 严胡勇.扩展粗糙集模型及其在水质富营养化评价中的应用研究[D].北京:中国科学院大学,2017.
- [11] PAWLAK Z. Rough sets[J]. International Journal of Parallel Programming,1982,11(5):341-356.
- [12] PAWLAK Z. Rough set theory and its applications to data analysis[J]. Cybernetics & Systems,1998,29(7):661-688.
- [13] MURATA T. Petri nets: properties, analysis and applications [J]. Proceedings of the IEEE,1989,77(4):541-580.
- [14] ZURAWSKI R,ZHOU Mengchu. Petri nets and industrial applications:a tutorial[J]. IEEE Transactions on Industrial Electronics,1994,41(6):567-583.
- [15] YAN Huyong,HUANG Yu,WANG Guoyin,et al. Water eutrophication evaluation based on rough set and petri nets:a case study in Xiangxi-River, Three Gorges Reservoir[J]. Ecological Indicators,2016,69:463-472.
- [16] 刘斌,岳会宇,李卓,等.支持向量机在油层含油识别中的应用[J].控制工程,2006,13(4):355-357.
- [17] 侯利娟,王国胤,聂能,等.粗糙集理论中的离散化问题[J].计算机科学,2000,27(12):89-94.
- [18] 于锐,刘知贵,黄正良.粗糙集理论应用中的离散化方法综述[J].西南科技大学学报:自然科学版,2005,20(4):32-36.
- [19] 张长胜,孙吉贵,欧阳丹彤.一种自适应离散粒子群算法及其应用研究[J].电子学报,2009,37(2):299-304.
- [20] 谢宏,程浩忠,牛东晓.基于信息熵的粗糙集连续属性离散化算法[J].计算机学报,2005,28(9):1570-1574.
- [21] 桑雨,李克秋,闫德勤.基于改进 χ^2 统计的数据离散化算法[J].大连理工大学学报,2012,52(3):443-447.
- [22] 赵静娴,倪春鹏,詹原瑞,等.一种高效的连续属性离散化算法[J].系统工程与电子技术,2009,31(1):195-199.