

引用格式:李昌,沈安江,常少英,等.机器学习法在碳酸盐岩岩相测井识别中应用及对比——以四川盆地MX地区龙王庙组地层为例[J].油气藏评价与开发,2021,11(4):586-596.

LI Chang, SHEN Anjiang, CHANG Shaoying, et al. Application and contrast of machine learning in carbonate lithofacies log identification: A case study of Longwangmiao Formation of MX area in Sichuan Basin[J]. Petroleum Reservoir Evaluation and Development, 2021, 11(4): 586-596.

DOI: 10.13809/j.cnki.cn32-1825/te.2021.04.015

机器学习法在碳酸盐岩岩相测井识别中应用及对比

——以四川盆地MX地区龙王庙组地层为例

李昌^{1,2}, 沈安江^{1,2}, 常少英^{1,2}, 梁正中³, 李振林⁴, 孟贺¹

(1. 中国石油杭州地质研究院, 浙江 杭州 310023; 2. 中国石油天然气集团公司碳酸盐岩储层重点实验室, 浙江 杭州 310023; 3. 榆林学院, 陕西 榆林 719000; 4. 中国石油测井集团公司, 陕西 西安 710077)

摘要:机器学习法是碳酸盐岩岩相测井识别的主要技术手段,针对不同地质情况和资料,选择适用的机器学习方法是高精度识别岩相的关键因素之一,然而针对机器学习识别岩相方法的适用性研究较少,为此列举了4种最常用的机器学习识别岩相方法:自组织神经网络聚类分析法(SOM)、基于图像多分辨率聚类分析法(MRGC)、K最近邻分类算法(KNN)和神经网络法(ANN)。通过对比方法的原理及方法的实际应用效果,总结出这4种机器学习法的优缺点和适用性,少量岩心样本情况下,优选MRGC法;在较多数量岩心资料情况下,优选KNN或MRGC法。其在四川盆地MX地区龙王庙组地层岩相识别应用中表明:MRGC和KNN法效果最好,其次为SOM法,ANN法效果最差。不同机器学习方法实际应用及对比分析成果,对于碳酸盐岩岩相测井识别方法在其他层组或其他工区的应用起到借鉴作用,并具有较强的实用价值。

关键词:碳酸盐岩岩相;测井识别;机器学习;龙王庙组;四川盆地

中图分类号:TE19

文献标识码:A

Application and contrast of machine learning in carbonate lithofacies log identification: A case study of Longwangmiao Formation of MX area in Sichuan Basin

LI Chang^{1,2}, SHEN Anjiang^{1,2}, CHANG Shaoying^{1,2}, LIANG Zhengzhong³, LI Zhenlin⁴, MENG He¹

(1. Hangzhou Research Institute of Geology, RIPED, CNPN, Hangzhou, Zhejiang 310023, China; 2. Key Laboratory of Carbonate Reservoirs, CNPN, Hangzhou, Zhejiang 310023, China; 3. Yulin University, Yulin, Shaanxi 719000, China; 4. China Petroleum Logging Co., Ltd., Xi'an, Shaanxi 710077, China)

Abstract: The machine learning method is the main technical means of carbonate lithofacies log identification. Selecting the appropriate machine learning method according to the different geological conditions and data is one of the key factors for high-precision identification of lithofacies. However, there are few researches on the applicability of machine learning identification methods. In this paper, four most commonly used machine learning methods for identifying lithofacies are studied, including Self Organizing Maps (SOM), Multi-Resolution Graph-based Clustering (MRGC), K Nearest Neighbor (KNN), and Artificial Neural Network (ANN). By comparing the principle and practical application effects of these methods, the advantages, disadvantages and applicability of the four machine learning methods have been summarized. When there are few core samples, MRGC is preferred, while when there are more core data, KNN is preferred as well as MRGC. Their application of lithofacies identification in the Longwangmiao Formation in the MX area in Sichuan Basin shows that MRGC and KNN are the best, SOM is the second, and ANN is the worst. This study of the application effects of machine learning methods provides a guidance for the identification of carbonate rock facies in other layers and regions, and has strong practical value.

Keywords: carbonate facies, logging identification, machine learning method, Longwangmiao Formation, Sichuan Basin

收稿日期:2020-09-09。

第一作者简介:李昌(1978—),男,硕士,高级工程师,从事碳酸盐岩岩相识别及储层孔隙结构测井评价工作。地址:浙江省杭州市西溪路920号中国石油杭州地质研究院,邮政编码:310023。E-mail:lic_hz@petrochina.com.cn

基金项目:国家科技重大专项“寒武系—中新元古界碳酸盐岩规模储层形成与分布研究”(2016ZX05004002);国家科技重大专项“四川盆地大型碳酸盐岩气田开发示范工程”(2016ZX05052);陕西省科技厅“神府地区煤系非常规天然气共生聚集机制及开发潜力评价”(2020SF-369)。

碳酸盐岩储层的发育与颗粒滩密切相关,储层发育受控于岩相分布^[1]。沉积微相精细研究对于储层预测非常重要。由于取心井资料有限,因此,利用测井资料准确识别岩相,对于颗粒滩的预测有着重要意义。由于碳酸盐岩储层受强烈成岩作用改造,裂缝、溶蚀孔洞发育,非均质性较强^[2-3],导致高精度识别基于岩石结构组分分类的岩相一直是个难题。机器学习方法效率高、速度快且准确率高,在岩心资料丰富的情况下,是目前岩相测井识别最主要的技术手段。机器学习方法较多,最常用的方法主要有聚类分析法^[4-6]、神经网络法^[7-10]和分类算法等等。针对不同地质情况和资料,合理选择适用的机器学习方法是高精度识别岩相的关键因素之一,然而针对机器学习方法的适用性分析较少,为此选择4种最常用的机器学习方法,即自组织神经网络聚类分析法(SOM)、基于图像多分辨率聚类分析法(MRGC)、K最近邻分类法(KNN)和神经网络法(ANN)。以四川盆地MX地区龙王庙组地层为例,选择2口岩心样本井及2口识别验证井,开展岩相测井识别效果对比分析研究,识别结果表明:MRGC识别精度最高,符合率在80%以上;其次为KNN识别,符合率在79%以上;再次为SOM识别,符合率在76%以上;ANN识别精度较低,符合率在62%以上。研究认为在少量岩心样本情况下,优选MRGC法;在较多数量岩心资料情况下,优选KNN或MRGC法。通过上述研究,总结了不同方法的优点及适用性,为机器学习方法在其他层组或其他地区开展碳酸盐岩岩相测井识别应用工作起到了借鉴作用,并具有较强的实用价值。

1 基本原理

1.1 自组织神经网络聚类分析(SOM)

自组织映射神经网络是一种无监督训练的神经网络,它通过引入邻域函数的概念,实现自组织无指导的学习,即,将所有神经元置于一个根据先验知识而事先确定的拓扑结构上去。邻域函数的引入使得拓扑结构限制SOM训练,这样可以最大限度地保障训练不会陷入局部最小^[11]。采用二维SOM结构,由输入层和竞争层构成。输入层的维数与输入样本向量维数是一致的,竞争层节点一般呈二维阵列分布,竞争层中的一个节点代表一个神经元,各神经元之间为侧抑制连接,输入层和竞争层之间采取全连

接^[12-16]。

SOM算法实现的具体步骤如下:

- 1) 网络初始化:初始化权值 W 等。
- 2) 从输入样本中随机选取一个输入向量 $X_i=(x_1, x_2, x_3, \dots, x_n)$ 。
- 3) 寻找竞争层获胜神经元:计算 X_i 与竞争层神经元 j 之间的距离,从中找出距离 X_i 最小的获胜神经元 g 。
- 4) 权值调整:按第一步调整神经元 g 和其邻域 $N(t)$ 内包含的神经元的权值。

$$W_{ij} = W_{ij} + \eta(X_i - W_{ij}) \quad (1)$$

式中: W_{ij} 表示输入层节点 i 和竞争层节点 j 的权值; η 表示学习率,一般随着进化次数而减小。

- 5) 判断算法是否结束,若未结束,返回第二步。

1.2 基于图像多分辨率聚类分析(MRGC)

基于图像多分辨率聚类分析方法(Multi-Resolution Graph-based Clustering, MRGC)是一种基于非参数K邻近法和图形数据表示的多维点阵图像识别方法。由于它不依靠判别类域来确定所属类别,而是靠周围有限邻近样本来分类,对于类域交叉或重叠较多的岩心样本集来说,MRGC方法较其他方法更为合适。

MRGC法的采样点与采样点之间的相似程度用欧氏距离来度量,根据欧式距离矩阵判断采样点与采样点之间的吸引与被吸引的关系。综合考虑每个采样点与其他采样点之间的吸附关系,每个采样点对其他所有采样点吸引的能力用近邻指数 NI 表示,将 NI 值最大的采样点作为最终的吸引中心。整个样本集被划分为多个吸引集,吸引集中心用核代表指数(KRI)表示,通过核代表指数的降序排列获取多层次分类中每一个层次的分类数,通过多层次的融合吸引集得到最终的分类结果^[17]。

MRGC采用向量空间模型,将优选的模型曲线和预测曲线转化为若干特征组成空间 $(t_1, t_2, t_3, \dots, t_k)$,再将各个特征样本在曲线数据上赋予的数值填充到向量空间中。最后一个样本数据 d_j 的数学表示形式为:

$$W_j(W_{1j}, W_{2j}, \dots, W_{kj}, \dots, W_{|T|j}) \quad (2)$$

式中: W_{kj} 表示特征 t_k 在样本数据 d_j 上的值; $|T|$ 表示特征向量的维数^[18]。

模型曲线和预测曲线样本之间的相似度利用

2个向量的几何关系来度量。设有2个特征向量 $X=(x_1, x_2, \dots, x_{|T|})$ 和 $Y=(y_1, y_2, \dots, y_{|T|})$, 则其相似度采用欧几里德距离公式计算,如下式:

$$\text{sim}(X, Y) = \frac{\sum_{i=1}^{|T|} x_i y_i}{\sqrt{\sum_{i=1}^{|T|} x_i^2} \cdot \sqrt{\sum_{i=1}^{|T|} y_i^2}} \quad (3)$$

通过式(3)可以计算待预测的特征向量与数据集的所有特征向量的距离,利用高斯函数将距离转换为权值,根据距离的远近对预测结果进行贡献值补偿,再通过每个最邻近特征样本乘以相应权重,然后将所有得到的结果累加,并除以所有权重值的和,如下式:

$$P = \frac{\sum_{i=1}^k S_i W_i}{\sum_{j=1}^k W_j} \quad (4)$$

式中: P 为最终的预测结果; S_i 为 k 个最邻近样本中的第 i 个特征样本; W_i 为 S_i 对应权重值。

1.3 K最近邻分类算法(KNN)

K最近邻分类算法(k Nearest Neighbor, KNN),是最简单的数学分类识别算法之一。每个样本都可以用它最接近的 k 个邻居来代表,根据数据集事先分类作为依据对新样本直接分类,无需学习和训练。如果一个待分样本的邻居为已经正确分类的对象,那么该待分样本所属的类别依据最邻近的一个或者几个样本的类别来决定。因此,KNN方法不受异常点影响,适合重叠较多的待分样本集或类域交叉的分类问题^[19-20]。KNN法算法简单、直接,在样本量少或样本特征少的情况下也可以分类,但要求样本类型的数量要均衡。

KNN算法实现的具体步骤如下:

假定有 c 个类别, k_1, k_2, \dots, k_c 的模式分类问题,每类有标明类别的样本 $N_i (i=1, 2, \dots, c)$ 个,规定 k_i 类的判别函数为

$$g_i(x) = \min \left\| x - x_i^{(k=1, 2, \dots, N)} \right\| \quad (5)$$

式中: i 为第 k_i 类; k 为 k_i 类 N_i 个样本中的第 k 个。

按照式(5)决策规则可以写为:若 $g_i(x) = \min g_i(x)$,则决策 $x \in k_i$ 。这一决策方法称为最近邻法,即对未知样本 x ,只要比较 x 与 $n = \sum_{i=1}^m n_i$ 个已知类别样本之间

的欧氏距离,决策 x 与离它最近的样本同类,就能判别样本 x 的类别。

1.4 神经网络方法(ANN)

神经网络(Artificial Neural Network, ANN)被广泛用于分类及识别领域。BP神经网络是最广泛采用的网络类型,包含输入层、隐含层和输出层,是一种按误差逆传播算法训练的多层前馈网络,其最主要的特点就是按照误差逆向传播法进行网络的学习与训练,其网络学习能存贮大量的输入和输出模式映射关系^[21]。

ANN算法实现的具体步骤如下:

1) 网络初始化

给各连接权值分别赋一个区间 $(-1, 1)$ 内的随机数,设定误差函数 e ,给定计算精度值 ϵ 和最大学习次数 M 。

2) 随机选取第 k 个输入样本及对期望输出

$$d_o(k) = [d_1(k), d_2(k), \dots, d_q(k)] \quad (6)$$

$$x(k) = [x_1(k), x_2(k), \dots, x_n(k)] \quad (7)$$

式(6)—(7)中: $d_o(k)$ 为输入层输入向量; $d_q(k)$ 为输入样本; q 为输入层神经元数; $x(k)$ 为期望输出向量; $x_n(k)$ 为期望输出值; n 为输出层神经元数。

3) 计算隐含层各神经元的输入和输出。

4) 利用网络期望输出和实际输出,计算误差函数对输出层的各神经元的偏导数 $\delta_o(k)$ 。

5) 利用隐含层到输出层的连接权值、输出层的 $\delta_o(k)$ 和隐含层的输出计算误差函数对隐含层各神经元的偏导数 $\delta_h(k)$ 。

6) 利用输出层各神经元的 $\delta_o(k)$ 和隐含层各神经元的输出来修正连接权值 $w_{ho}(k)$ 。

7) 利用隐含层各神经元的 $\delta_h(k)$ 和输入层各神经元的输入修正连接权。

8) 计算全局误差

$$E = \frac{1}{2m} \sum_{k=1}^m \sum_{o=1}^q [d_o(k) - y_o(k)]^2 \quad (8)$$

式中: E 为全局误差; m 为样本个数; $y_o(k)$ 为输出层输出向量。

9) 判断网络误差是否满足要求。当误差达到预设精度或学习次数大于设定的最大次数时,则结束算法。否则,选取下一个学习样本及对应的期望输出,返回到第三步,进入下一轮学习。

1.5 测井相分析方法

测井相是能反映沉积物沉积特征,并能使这个沉积物与其他沉积物区别开的一组测井特征及其组合。岩相是指形成岩石所处的地质环境的总和,包括温度、气候、地层、时代等。是一定沉积环境中形成的岩石或岩石组合。每类岩相具有不同物理性质,例如高孔或低孔,高放射性或低放射性,含油气或含水和裂缝是否发育等等。导致同一岩相可能具有多种测井特征,因此,一种岩相可能对应多个测井相。

测井相个数可以根据实际需要设定,一般测井相个数要大于岩相个数,测井相个数越多则与岩相对应关系也越复杂。据实际生产经验,一般测井相个数为岩相个数的1~3倍比较合适。研究区岩相划分为3类,测井相个数设定为9类。

测井相分析方法就是建立测井相与岩相对应关系,将测井相转化为岩相,从而实现岩相识别。测井相与岩相对应关系是基于岩心标定建立的。由于岩—电关系的复杂性,测井相与岩相往往不能完全一一对应,需要参考每种测井相对应不同岩相的概率大小来确定对应关系,概率等于每种测井相对应不同岩相的累计厚度占该测井相总厚度的百分比。以MX12井SOM方法为例,测井相1总厚度为9 m,测井相1对应颗粒云岩部分的厚度为7.25 m,则测井相1对应颗粒云岩相概率为86.4%,同理,对应粉晶云岩相概率为13.6%,对应泥晶云岩相概率为0%,综合

认为测井相1对应颗粒云岩相(表1)。同样基于岩心标定概率百分比,颗粒云岩相对应测井相1、测井相4和测井相7,其他岩相类似选择概率大于70%的,建立测井相与岩相对应关系。

2 技术应用

2.1 岩石学特征

四川盆地中部MX地区龙王庙组呈北东—南西向展布,岩性主要由晶粒白云岩、泥晶白云岩及颗粒白云岩组成^[22-23](图1)。

龙王庙组岩石类型主要有3种:

1) 颗粒云岩:主要为砾屑云岩和鲕粒云岩(图2d),少量豆粒云岩。发育在颗粒滩的主体部位,对应水体能量较高的沉积环境,溶蚀孔洞发育,是龙王庙组最主要的储集层岩石类型。

2) 晶粒云岩:以粉晶云岩和细晶白云岩为主(图2a、图2c),少量中晶云岩。晶粒白云岩主要为部分颗粒白云岩经重结晶改造而形成晶粒白云岩的储集空间,以晶间孔和不规则的溶蚀孔洞为主。

3) 泥晶云岩:包括泥晶云岩(图2b)和泥质泥晶云岩。一般发育在低能局限静水环境。

地质上岩性分类较多,岩性在纵向上变化较快,而测井曲线受分辨率所限,不能反映薄层岩性信息。另外,岩性类型过多,也会导致岩性识别精度降低。岩性识别目的是利用岩性信息分析沉积时期水体能量,从而满足沉积微相精细研究需求。因此,综

表1 测井相与岩心岩相标定概率百分比
Table 1 Calibration probability percentage of logging facies and core facies

SOM法 测井相类型	岩心标定概率百分比			MRGC法 测井相类型	岩心标定概率百分比		
	颗粒云岩	粉晶云岩	泥晶云岩		颗粒云岩	粉晶云岩	泥晶云岩
测井相1	86.4	13.6	0	测井相1	80.6	19.4	0
测井相2	0	100.0	0	测井相2	0	100.0	0
测井相3	0	0	100.0	测井相3	100.0	0	0
测井相4	100.0	0	0	测井相4	0	100.0	0
测井相5	0	93.3	6.7	测井相5	84.6	15.4	0
测井相6	0	100	0	测井相6	0	0	100.0
测井相7	95.5	0	4.5	测井相7	11.8	70.6	17.6
测井相8	16.7	83.3	0	测井相8	73.7	26.3	0
测井相9	0	0	100.0	测井相9	0	0	100.0

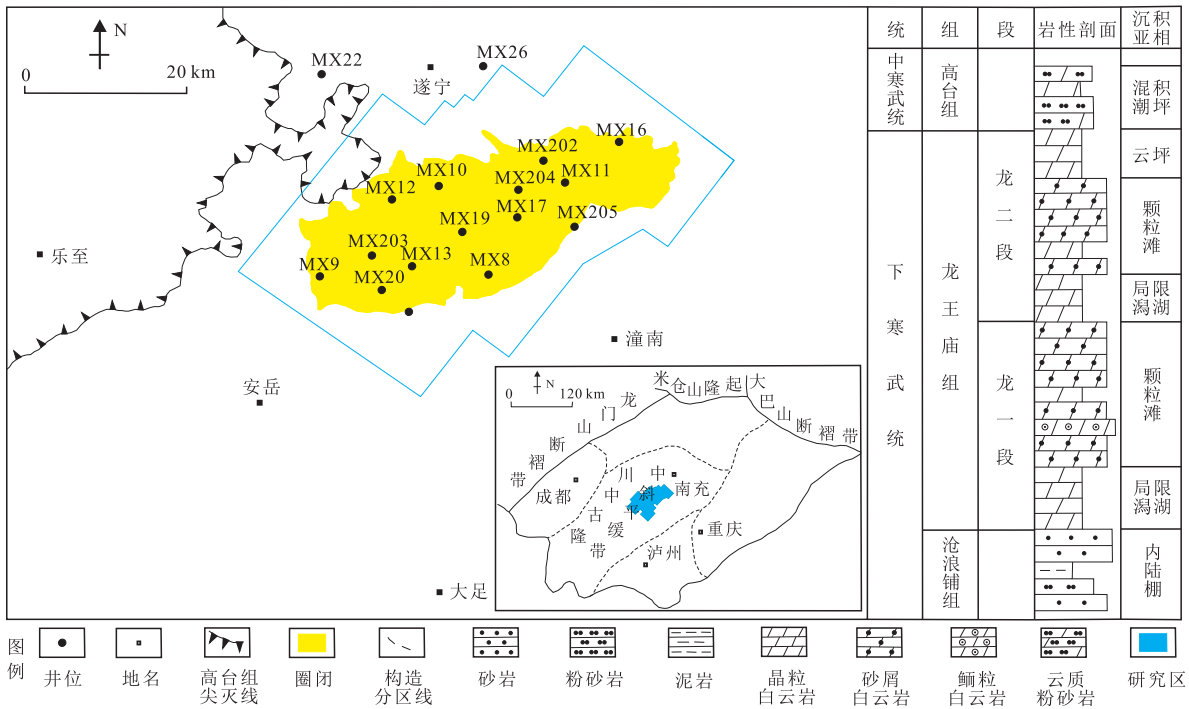


图1 研究区位置、构造、地层剖面(据司马立强^[3])

Fig. 1 Location, structure and stratigraphic section of study area(According to SIMA Liqiang^[3])

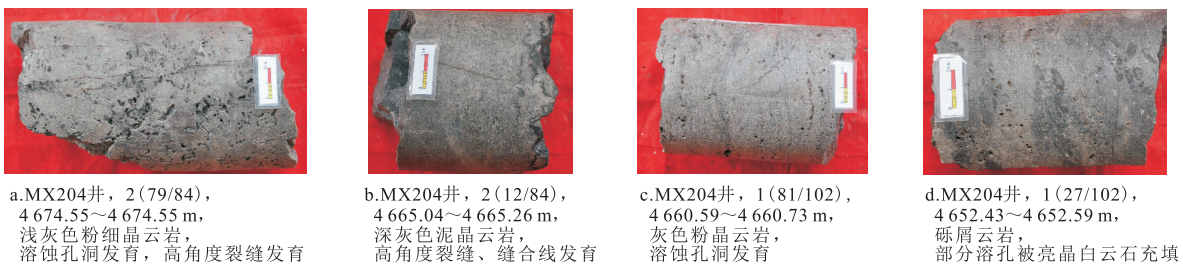


图2 MX地区龙王庙组岩石学特征及沉积构造特征

Fig. 2 Petrological and sedimentary structural characteristics of Longwangmiao Formation in MX area

合考虑沉积微相研究需求,同时兼顾识别精度,将岩相划分为3大类(表2)。由于颗粒云岩和细晶云岩粒径和晶体尺寸都比较粗,在测井特征上难以区分,而细晶云岩和颗粒云岩都反映高能沉积环境,因此,统一归为颗粒云岩相。颗粒云岩相代表高能沉积环

境,粉晶云岩相代表中等沉积环境,泥晶云岩相则代表低能沉积环境,划分为3大类岩相,既能满足生产需求,也保证了较高精度的岩相识别。

2.2 岩相—测井特征分析

基于上述岩相划分,分析岩相与测井特征关系,由图3可知测井数据重叠性很强,不同岩相的测井特征总结如下。

1) 颗粒云岩相:测井特征为低自然伽马,分为致密和溶孔发育2种情况。在致密情况下,测井特征为低时差和高电阻率;当溶孔发育时,表现为高声波时差,电阻率相对降低;如果为纯水层,则电阻率为更低值,声波时差为更高值。

表2 龙王庙组岩相分类
Table 2 Lithofacies classification of Longwangmiao Formation

岩相	岩性
颗粒云岩相	颗粒云岩,砾屑云岩、砂屑云岩、细晶云岩等粗结构的白云岩
粉晶云岩相	粉晶云岩、细—粉晶云岩等粉晶级别的白云岩
泥晶云岩相	泥晶云岩,泥质泥晶云岩,泥质泥—粉晶云岩等

2) 粉晶云岩相:测井特征与颗粒云岩相似,为低自然伽马,物性上较颗粒云岩相略差,其声波时差略低于细晶云岩,声波时差值为中等,电阻率相对降低;当为纯水层时,则电阻率更低,声波时差更高;当为致密不发育孔隙时,测井特征为高电阻率,低声波时差。

3) 泥晶云岩相:分为高泥含量和低泥含量2种情况,泥含量低时,孔隙不发育,表现为中等自然伽马、低声波时差和高电阻率特征;当含有较高泥质含量时,测井特征显著,高自然伽马、高声波时差、低电阻率特征。

2.3 应用效果

选择2口取心井(MX12井、MX203井)作为岩心样本井,另外选择2口取心井(MX17井、MX202井)作为识别效果验证井。输入测井曲线选择声波时差、密度、中子、深测向电阻率和自然伽马。分别采用SOM、KNN、MRGC和ANN方法进行识别并对比识别效果。

2.3.1 SOM方法和MRGC方法识别效果

SOM和MRGC都属于聚类分析方法。聚类分析法识别岩相过程是首先确定聚类测井相个数,然后聚类获得测井相,接着与取心井进行岩心标定,建立测井相与岩相的对应关系,最后将测井相转换为岩相,实现非取心井和非取心段的岩相识别。测井相

和岩相对应关系是决定识别效果的关键,可以基于大量岩心对比总结的经验,也可以根据已有测井理论基础来确定对应关系,因此,即使岩心样本数量很少,聚类分析法也可以识别岩相。

聚类分析的关键是测井相分析,即测井相转化为地质相,转化过程可能会出现一种岩相对应多个测井相,例如从图4中可以看出,颗粒云岩在中子、密度和声波时差测井特征上具有变化,说明孔隙度具有高低变化。以SOM法为例,高孔隙度颗粒云岩对应测井相1,中等孔隙度颗粒云岩对应测井相4,低孔隙度颗粒云岩对应测井相7。

基于岩相和测井相对应关系(图5),开展对验证井(MX17井、MX202井)岩相识别,识别效果见图6和图7。SOM方法下,MX17井取心段厚度为68.8 m,岩相识别符合厚度为52.8 m,识别符合率为76.7%;MX202井取心段厚度为62.2 m,岩相识别符合厚度为49.8 m,识别符合率为80%。MRGC方法下,MX17井取心段厚度为68.8 m,岩相识别符合厚度为56.3 m,识别符合率为81.8%;MX202井取心段厚度为62.2 m,岩相识别符合厚度为50 m,识别符合率为80.4%。

2.3.2 KNN和ANN方法识别效果

KNN和ANN法都需要确定的岩心学习样本,KNN是一种强制性分类方法,严格将已知岩心样本作为标准,而ANN是基于学习样本进行训练,获取人

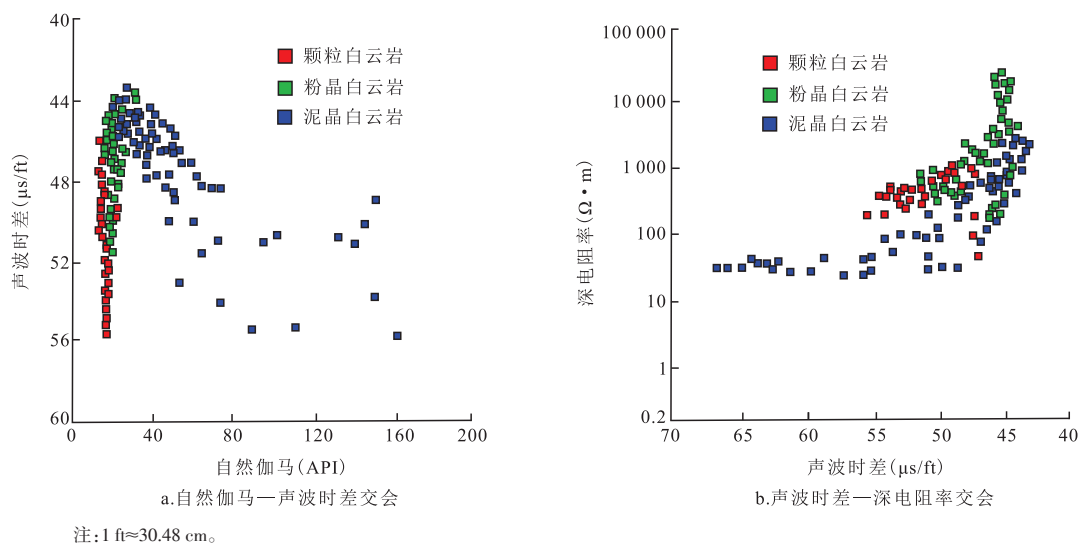


图3 不同岩相测井特征交会图
Fig. 3 Cross plot of logging characteristics of different lithofacies

工智能,因此,岩心样本的数量和典型性非常重要。根据部分岩心样本数据(表3),不同岩相类型的样本数量基本均衡,其中泥晶白云岩相7个,累计厚度为31.83 m;粉晶白云岩相岩样13个,累计厚度为29.73 m;颗粒白云岩相8个,累计厚度为36.66 m。

开展MX17井和MX202井岩相识别,识别结果见图6和图7。KNN识别效果较好,MX17井取心段厚度为68.8 m,岩相识别符合厚度为54.7 m,识别符合率为79.5%;MX202井取心段厚度为62.2 m,岩相

识别符合厚度为50.2 m,识别符合率为80.7%。ANN识别效果较差,MX17井取心段厚度为68.8 m,岩相识别符合厚度为47.3 m,识别符合率为68.7%;MX202井取心段厚度为62.2 m,岩相识别符合厚度为38.8 m,识别符合率为62.3%。

2.4 不同方法适用性分析

不同的算法各有优缺点、适用性及应用条件。另外测井资料的品质要好,在井径扩径,密度易失真

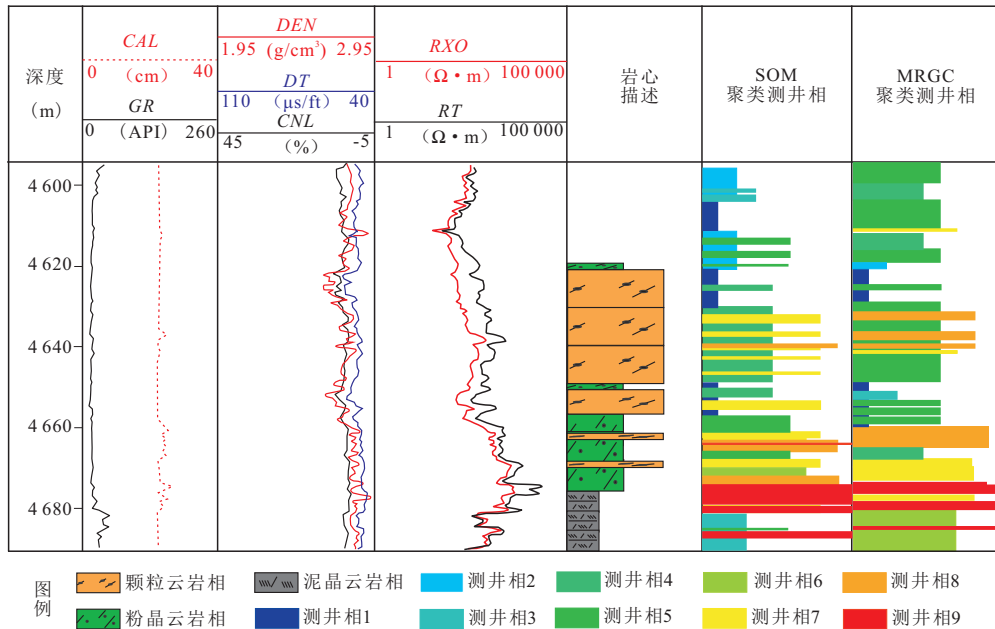


图4 SOM和MERGC聚类测井相与岩心标定(MX12井)

Fig. 4 Cluster logging facies and core calibration by SOM and MERGC method (Well-MX12)

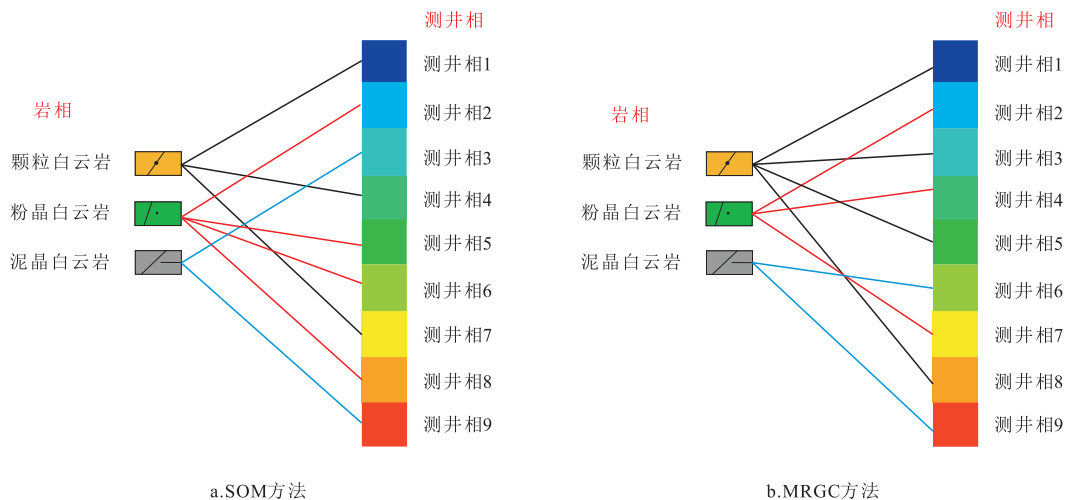


图5 岩相与测井相对应关系

Fig. 5 Correlation between lithofacies and logging

的情况下,尽量不使用密度曲线。

MRGC和KNN方法都是适用于类域交叉或重叠较多待分样本的情况,对于碳酸盐岩岩相,其测井特征在二维交会图上往往都是交叉和重叠较多,因此,更适合测井进行岩相分类。不同的是KNN并不对样本聚类,而是直接参照已有岩心样本的岩相类型分类,获取不同岩相类型的测井特征,以这些特征为标准,对测井曲线进行分类,并确定对应的岩相类型。因此,要求岩心样本岩相类型数量均衡和测井特征要准确。MRGC是对已有岩心样本聚类进行优化,获取优化后的岩心样本分类作为分类依据,其优点是可以消除一些薄层及不符合测井原理的异常点。

SOM法是通过神经网络训练优化样本的,其优点是稳定性好,岩心样本数量较多情况下,也能获得很好效果,但由于碳酸盐岩取心数据往往有限,因此,MRGC法和KNN法要优于SOM法和ANN法。

对4种方法优缺点和适用性总结见表4,在实际应用中,可以根据地质情况及测井资料情况选择适用的方法。

3 结论

以四川盆地MX地区龙王庙组地层为例,通过对4种机器学习方法岩相测井识别实际应用效果的对比

表3 取心井岩心岩相样本(部分)
Table 3 Lithofacies of core samples(part)

序号	岩相	厚度(m)	声波时差(AC) ($\mu\text{s}/\text{ft}$)	自然伽马(GR) (API)	中子(NPHI) (V/V)	密度(DEN) (g/cm^3)	深电阻率(RT) ($\Omega\cdot\text{m}$)
1	颗粒云岩	4.25	48.24	17.71	0.082	2.77	103.59
2	颗粒云岩	5.49	47.92	19.10	0.091	2.79	51.91
3	颗粒云岩	1.24	49.05	17.63	0.064	2.74	425.84
4	颗粒云岩	16.99	49.35	17.60	0.052	2.73	809.46
5	颗粒云岩	5.91	50.08	18.04	0.061	2.70	595.41
6	颗粒云岩	0.93	47.98	20.53	0.050	2.75	818.09
7	颗粒云岩	0.81	47.46	19.45	0.041	2.83	1 649.04
8	颗粒云岩	1.04	45.08	19.17	0.034	2.83	4 071.69
9	粉晶云岩	1.17	43.62	32.17	0.036	2.87	1 036.42
10	粉晶云岩	4.75	45.91	18.58	0.070	2.81	381.87
11	粉晶云岩	1.76	45.08	28.45	0.054	2.82	777.23
12	粉晶云岩	0.70	44.14	22.85	0.041	2.82	1 258.25
13	粉晶云岩	0.70	44.22	26.39	0.030	2.87	13 848.18
14	粉晶云岩	0.58	44.06	16.49	0.035	2.83	8 054.69
15	粉晶云岩	0.81	43.86	19.15	0.035	2.80	9 937.23
16	粉晶云岩	0.73	48.18	18.53	0.047	2.79	506.02
17	粉晶云岩	1.97	50.59	18.58	0.055	2.72	726.43
18	粉晶云岩	1.85	49.47	18.28	0.064	2.72	469.56
19	粉晶云岩	2.55	47.49	19.94	0.040	2.80	2 024.54
20	粉晶云岩	5.56	46.51	19.49	0.041	2.81	1 858.67
21	粉晶云岩	6.60	45.57	18.94	0.036	2.81	7 662.91
22	泥晶云岩	2.00	46.45	25.89	0.060	2.80	210.82
23	泥晶云岩	3.89	44.62	26.17	0.045	2.84	530.76
24	泥晶云岩	2.66	43.65	22.57	0.032	2.86	10 060.70
25	泥晶云岩	3.36	43.77	22.41	0.038	2.83	7 761.66
26	泥晶云岩	4.29	43.49	20.58	0.029	2.83	12 356.61
27	泥晶云岩	1.85	43.64	19.02	0.027	2.86	7 996.26
28	泥晶云岩	13.78	46.15	32.67	0.042	2.82	2 495.68

注:1 ft \approx 30.48 cm。

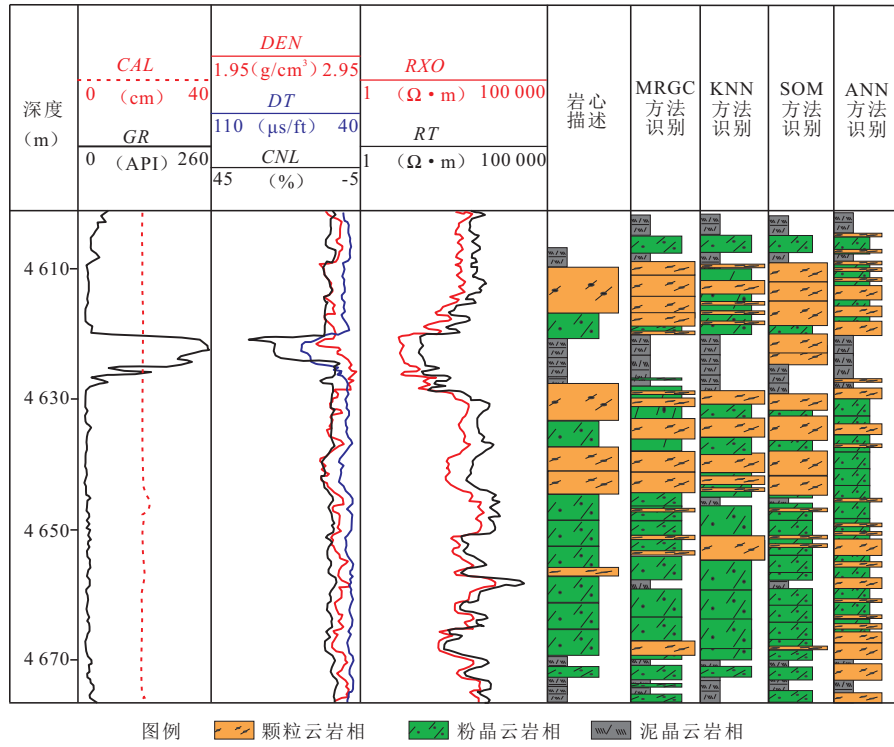


图6 4种方法岩相识别结果对比(MX17井)

Fig. 6 Comparison of lithofacies identification results of four methods (Well-MX17)

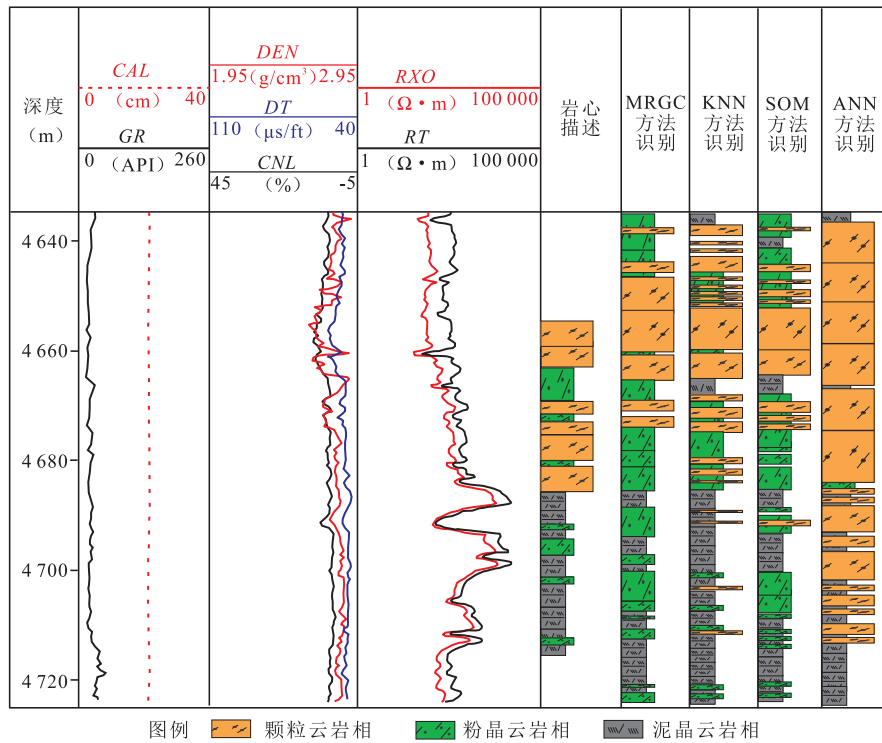


图7 4种方法岩相识别结果对比(MX202井)

Fig. 7 Comparison of lithofacies identification results of four methods (Well-MX202)

表4 不同识别方法优缺点总结
Table 4 Advantages and disadvantages of different identification methods

方法	优缺点		应用条件		方法关键点
	优点	缺点	测井资料要求	岩心资料要求	
自组织映射神经网络聚类分析(SOM)	容错性强,稳定性好,具有自联想性	人工经验主导聚类测井相数目,步骤较多			岩相和测井相对应关系
基于图像多分辨率聚类分析(MRGC)	善于解决类域交叉或重叠较多的问题	人工经验主导聚类测井相数目,人为影响因素大,步骤多	至少5项常规测井资料(中子、密度、声波、自然伽马、电阻率)	少量岩心资料	
K最近邻分类算法(KNN)	善于解决类域交叉或重叠较多的问题,无需训练,方法简单	样本数量不均衡或较少时,预测偏差大		中等数量岩心资料,且每类岩相类型样本数量均衡	不同类型样本数量要均衡
神经网络方法(ANN)	较强容错能力及自学习自适应能力	容易陷入局部极小化问题,对样本数量依赖性强		大量岩心样本资料	样本数据量大而且要具有代表性

比分析,总结出这4种方法的适用性,具体如下:

1) 对于无监督学习的聚类分析方法,如SOM和MRGC法,关键在于建立岩相与测井相之间的对应关系,其中SOM法容错性更强,MRGC法善于解决类域交叉问题,在岩相分类较多情况下,MRGC法效果更好。

2) 对于有监督学习的机器学习方法,如ANN和KNN法,关键在于学习训练样本的典型性和数量均衡性,ANN法自适应能力强,KNN法善于解决类域交叉问题,岩心样本数量在中等以下时,KNN法效果更好。

3) 在实际生产应用中,碳酸盐岩地层岩—电关系复杂且岩心数量有限,不同岩相的测井特征交叉重叠而难以区分,因此,KNN和MRGC法更适用。

参 考 文 献

[1] 沈安江,赵文智,胡安平,等.海相碳酸盐岩储集层发育主控因素[J].石油勘探与开发,2015,42(5):545-554.
SHEN Anjiang, ZHAO Wenzhi, HU Anping, et al. Major factors controlling the development of marine carbonate reservoirs[J]. Petroleum Exploration and Development, 2015, 42 (5): 545-554.

[2] 周进高,徐春春,姚根顺,等.四川盆地寒武统龙王庙组储集层形成与演化[J].石油勘探与开发,2015,42(2):158-166.
ZHOU Jingao, XU Chunchun, YAO Genshun, et al. Genesis and evolution of Lower Cambrian Longwangmiao Formation reservoirs, Sichuan Basin, SW China[J]. Petroleum Exploration and Development, 2015, 42(2): 158-166.

[3] 司马立强,陈志强,王亮,等.基于滩控岩溶型白云岩储层分

类的渗透率建模方法研究——以川中磨溪—高石梯地区龙王庙组为例[J].岩性油气藏,2017,29(3):92-102.

SIMA Liqiang, CHEN Zhiqiang, WANG Liang, et al. Permeability modeling based on the classification of beach-controlled karst dolomite reservoirs:a case from Longwangmiao Formation in Moxi-Gaoshiti area, central Sichuan Basin[J]. Lithologic Reservoirs, 2017, 29(3): 92-102.

[4] YE S J, RABLILLER P. A new tool for electro-facies analysis: multi-resolution graph-based clustering[C]// Paper SPWLA-2000-PP presented at the SPWLA 41st Annual Logging Symposium, Dallas, Texas, June 2000.

[5] 张振城,孙建孟,马建海,等.利用测井资料自动识别藻灰岩[J].吉林大学学报(地球科学版),2005,32(3):382-388.

ZHANG Zhencheng, SUN Jianmeng, MA Jianhai, et al. Automatic identification of algae limestone with well logging data[J]. Journal of Jilin University (Earth Science Edition), 2005, 32(3): 382-388.

[6] AGHCHELOU M, SHAHVAR M B. Lithofacies estimation by multi-resolution graph-based clustering of petrophysical well logs: Case study of South Pars Gas Field of Iran[C]// Paper SPE-162991-MS presented at the Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, August 2012.

[7] 于建华.应用神经网络自动识别岩性[J].石油地球物理勘探,1993,28(1):59-66.

YU Jianhua. Automatic lithology identification using artificial neural network[J]. Oil Geophysical Prospecting, 1993, 28 (1): 59-66.

[8] ALPANA B, HANS B H. Determination of facies from well logs using modular neural networks[J]. Petroleum Geoscience, 2002, 8: 217-228.

[9] HONG T, NIALL T, MEDDAUGH W S. Successful carbonate well log facies prediction using an artificial neural network

- method: Wafra Maastrichtian reservoir, partitioned neutral zone (PNZ), Saudi Arabia and Kuwait[C]// Paper SPE-123988-MS presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, October 2009.
- [10] HASSAN M, MOOSA E, ELHAM H A. Application of artificial neural networks for prediction of Sarvak Formation lithofacies based on well log data, Marun oil field, SW Iran[J]. *Geopersia*, 2015, 5(2): 111-123.
- [11] 桑应宾. 基于K近邻的分类算法研究[D]. 重庆: 重庆大学, 2009.
- SANG Yingbin. Research on classification algorithm based on K-nearest neighbor[D]. Chongqing: Chongqing University, 2009.
- [12] 周少华, 付略, 梁宝鏊. 基于SOM神经网络的古代青瓷聚类分析[J]. *中国科学E辑*, 2008, 38(7): 1089-1096.
- ZHOU Shaohua, FU Lue, LIANG Baoli. Clustering analysis of Ancient Celadon based on SOM neural network[J]. *Science in China(Series E)*, 2008, 38(7): 1089-1096.
- [13] 付小青, 张爱明. 基于SOM的入侵检测算法的特征选择[J]. *华中科技大学学报(自然科学版)*, 2007, 35(7): 5-7.
- FU Xiaoqing, ZHANG Aiming. Feature selection of SOM-based intrusion detection algorithm[J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2007, 35(7): 5-7.
- [14] KOHONEN T. Self-organized formation of topologically correct feature maps[J]. *Biological Cybernetics*, 1982, 43: 59-69.
- [15] CHANGA H C, KOPASKA- MERKELB D C, CHEN H C. Identification of lithofacies using Kohonen self-organizing maps[J]. *Computers & Geosciences*, 2002, 28: 223-229.
- [16] STUNDNER M, OBERWINKLER C. Self-Organizing maps for lithofacies identification and permeability prediction[C]// Paper SPE- 90720- MS presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, September 2004.
- [17] 张猛. 地层岩性的多层次自组织自动分类方法研究[D]. 武汉: 长江大学, 2018.
- ZHANG Meng. Study on multi-level self-organization automatic classification method for stratigraphy lithology[D]. Wuhan: Yangtze University, 2018
- [18] 吴进波. MRGC方法在储层分类及渗透率精细评价中的应用[J]. *海洋石油*, 2017, 37(1): 49-53.
- WU Jinbo. Application of reservoir classification and permeability elaborate evaluation with MRGC method[J]. *Offshore Oil*, 2017, 37(1): 49-53.
- [19] 寇莎莎, 魏振军. K-最近邻的改进及其在文本分类中的应用[J]. *河南师范大学学报(自然科学版)*, 2005, 33(3): 134-136.
- KOU Shasha, WEI Zhenjun. Improvement of K nearest neighbors and applications in text classification[J]. *Journal of Henan Normal University (Natural Science Edition)*, 2005, 33(3): 134-136.
- [20] 孙岩, 吕世聘, 王秀坤, 等. 基于结构学习的KNN分类算法[J]. *计算机科学*, 2007, 34(12): 184-186.
- SUN Yan, LYU Shiping, WANG Xiukun, et al. K- nearest neighbor algorithm based on learning structure[J]. *Computer Science*, 2007, 34(12): 184-186.
- [21] 张涛. 用神经网络研究元素与岩浆岩、沉积岩岩性及沉积岩矿物的关系[D]. 西安: 西北大学, 2016.
- ZHANG Tao. The relationships between rock elements and the igneous rocks, the lithologic discrimination and mineral identification of sedimentary rocks: A study based on the method of artificial neural network[D]. Xi'an: Northwest University, 2016.
- [22] 金民东, 曾伟, 谭秀成, 等. 四川磨溪—高石梯地区龙王庙组滩控岩溶型储集层特征及控制因素[J]. *石油勘探与开发*, 2014, 41(6): 650-660.
- JIN Mindong, ZENG Wei, TAN Xiucheng, et al. Characteristics and controlling factors of beach-controlled karst reservoirs in Cambrian Longwangmiao Formation[J]. *Petroleum Exploration and Development*, 2014, 41(6): 650-660.
- [23] 张建勇, 罗文军, 周进高, 等. 四川盆地安岳特大型气田下寒武统龙王庙组优质储层形成的主控因素[J]. *天然气地球科学*, 2015, 26(11): 2063-2074.
- ZHANG Jianyong, LUO Wenjun, ZHOU Jingao, et al. Main origins of high quality reservoir of Lower Cambrian Longwangmiao Formation in the Giant Anyue Gas Field, Sichuan Basin, SW China[J]. *Natural Gas Geoscience*, 2015, 26(11): 2063-2074.

(编辑 李青)