

基于密度聚类的 K 近邻法在储层流体识别中的应用

赵 军¹, 卢一凡¹, 李宗杰², 柳建华²

(1. 西南石油大学地球科学与技术学院, 四川成都 610500; 2. 中石化西北油田分公司, 新疆乌鲁木齐 830013)

摘要:针对传统储层流体识别方法识别精度低、运算量大、过于依赖个人经验的缺点,提出基于密度聚类的 K 近邻法,根据待测层段测井数据的空间分布规律,将样本按相对密度聚类成数据簇,并利用 K 近邻投票获得各簇所属类别。将该方法应用在某油田奥陶系鹰山组碳酸盐岩储层识别中。结果表明,较之其他常用识别方法,该算法识别精度高,泛化性和鲁棒性强,在处理大数据分类问题时具有明显优势,且在识别常规方法难以识别的油水同层时取得了较好的效果,具有良好的应用前景,为利用数据挖掘方法解决油田勘探开发中的复杂问题提供了新思路。

关键词:测井解释; 流体识别; K 近邻法; 相对密度聚类; 数据挖掘

中图分类号:P 631.84; TE 122.2 **文献标志码:**A

引用格式:赵军, 卢一凡, 李宗杰, 等. 基于密度聚类的 K 近邻法在储层流体识别中的应用[J]. 中国石油大学学报: 自然科学版, 2015, 39(5): 65-71.

ZHAO Jun, LU Yifan, LI Zongjie, et al. Application of density clustering based K -nearest neighbor method for fluid identification [J]. Journal of China University of Petroleum (Edition of Natural Science), 2015, 39(5): 65-71.

Application of density clustering based K -nearest neighbor method for fluid identification

ZHAO Jun¹, LU Yifan¹, LI Zongjie², LIU Jianhua²

(1. School of Geoscience and Technology, Southwest Petroleum University, Chengdu 610500, China;

2. Sinopec Northwest Oilfield Branch, Urumqi 830013, China)

Abstract: Reservoir fluid identification is an indispensable link in logging interpretation. In order to remove the defects of traditional approaches, such as unsatisfying accuracy, excessive computation, undue dependence on personal experience, a density clustering based K -nearest neighbor method was proposed. According to the spatial distribution of the interval logging data under test, data clusters are formed based on relative density. And then with K -nearest neighbor voting method, the categories of all clusters become available. Comparing with other commonly used identification methods, tested on the carbonate reservoir of Ordovician Yingshan Formation in an oil field, this approach shows a high accuracy, strong generalization and robustness, as well as better effects on oil-water layer identification which is usually difficult for the compared methods. The method has a good application prospect and provides a new thought on solving complex problems in oilfield exploration and development with data mining methods.

Keywords: logging interpretation; fluid identification; K -nearest neighbor method; relative density clustering; data mining

利用测井资料对储层流体进行精确识别是储层评价工作的重要任务。常规测井解释中,流体识别一般采用曲线重叠法^[1-2]和交会图法^[2-4],识别依据为测井曲线重叠时线间距的变化或测井数据交会分

析建立的图版,识别结果过于依赖经验判断,分类标准因人而异,精度难以保证,而目前对上述方法的改进仅限于判别因子与阈值的优化^[5-8],并未从实质上解决这一问题。现有的机器学习法,如 Fisher 判别

法^[9-11]、神经网络^[12-15]等,亦存在一定缺陷:前者依据组间与组内均方差之比最大的原则,寻找将多维数据点投影为一维数值的线性函数建立图版进行识别,但由于地下结构复杂,测井数据往往线性不可分,此时算法精度难以保证;后者通过某种算法构建参数间的映射模型^[15]以实现流体识别,但该算法学习效率固定,网络参数由个人经验或反复试验确定而缺乏理论指导,处理速度慢,且结果值得商榷。笔者根据测井数据结构,利用局部相对密度将待测样本点聚类成簇,并利用 K 近邻投票获得各簇所属类别。

1 原理方法

基于密度聚类的 K 近邻法(density clustering based K -nearest neighbor method, DCBKNN),是对相对密度聚类和 K 近邻法的综合,其主要思想是,利用相对密度聚类将待测样本中杂乱分布的数据点简化成为数据簇,之后只需利用 K 近邻投票判断每个簇的核心点所属的类别即可实现储层流体类型的识别。

1.1 相对密度聚类算法构建数据簇

聚类分析(clustering analysis)是一种根据数据结构特征的相似性将数据集划分成若干小组的数理分析方法,其中每个数据组均是由一系列隶属于同一类别的数据点构成的集合,称为数据簇(data cluster)^[16-19]。

基于密度的聚类方法认为整个数据空间由若干分布稠密的数据区域和零星分布的数据点构成,这些高密度区域构成的数据簇被分散数据点构成的稀疏空间所分隔,而算法的目的就是根据全局密度的差异找出这些数据簇。

常用的密度聚类法中最典型的是DBSCAN算法^[18-19],其核心参数为邻域半径 Eps 和邻域内最少元素数量 $minPts$ 。通过检索找出若干个高密度区域的核心点并扩展形成数据簇,要求簇内每个元素在半径 Eps 内都能至少包含 $minPts$ 个点。该方法存在一些难以克服的缺点。首先,聚类参数 Eps 和 $minPts$ 的敏感性过强,即使是细微的改变也能造成巨大的偏差。图1显示了不同的聚类参数对聚类结果的影响。当 Eps 和 $minPts$ 的设置合理时,图中分布紧密的数据能准确地聚类成两个簇,如图1(a)所示。此时略增大 $minPts$,使簇II中元素数量达不到 $minPts$ 的要求,机器会对聚类结果进行自动修改,造成将簇I和簇II误分成一类的结果,如图1(b)所示。其次,聚类参数的选择往往基于个人经验,且根据问题的不同,取值也往往大相径庭,很难给出一个完善的参考区间,这在处理一些复杂问题时给使用者造成了极大的不便。最后,该算法在核心点的提取和扩展过程中,繁冗的计算和储存过程会大量占用内存^[18]。可见,该方法在处理数据量大、结构复杂的储层流体识别问题时,实用价值很有限。

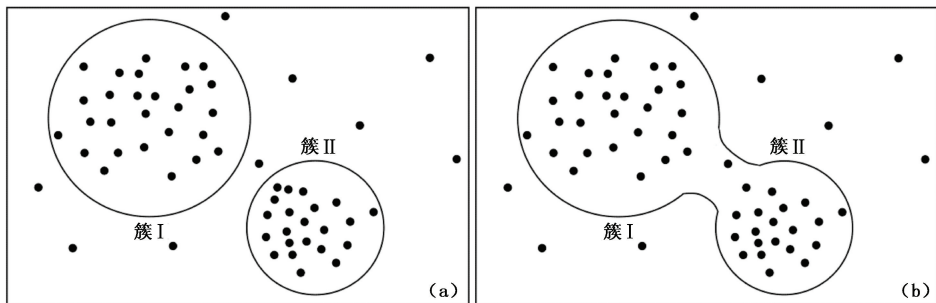


图1 不同的聚类参数对聚类结果的影响

Fig. 1 Distinct influence to clustering result when selecting different parameters

为了弥补DBSCAN的上述缺点,本文中引入 K 近邻相对密度以替换全局密度,不再使用 Eps 和 $minPts$ 这两个参数,而是通过设置阈值对簇的范围进行控制,按照相对密度的高低,将分布相对疏松的点排除出簇,达到了较好的聚类效果。

对于储层流体识别问题,首先需选择适当的测井识别因子。测井解释中一般选择GR(自然伽马)、RD(深侧向电阻率)、RS(浅侧向电阻率)、AC

(声波时差)、DEN(密度)、CNL(补偿中子)、SP(自然电位)等常规测井参数及其衍生参数作为识别因子。根据测试资料,提取流体类型已知的数据点构建参考样本集,提取待识别的数据点构成待测样本集。由于识别因子的取值范围差异较大,有时甚至相差多个数量级,因此需对参考样本集和待测样本集采取标准化处理以消除样本量纲不同造成的误差。

其次,选择一个合适的 K 值,对标准化后的待测数据集中所有元素依次检索其 K 个最近邻,并计算其平均 K 近邻距离 $nnad(x_i)$,选择出 $nnad(x_i)$ 最小的点作为核心点,将核心点与其 K 近邻作为初始簇。

对于 K 值的选择,可以通过选择一个较小的初值,再根据参考样本集进行交叉验证得到最优值^[20]。具体操作为,从参考样本集中随机抽取部分数据作为训练集而将剩余数据作为测试集,根据训练集元素的类别和分布确定初始 K 值,在测试集中进行验证,并根据需要对其进行适当修改,多次重复交叉验证过程,以得到最优 K 值。为了保证 K 近邻投票结果不会出现“平票”的情况,应尽量保证 K 为奇数。

数据集中某一点 x^* 与其 K 近邻集合 $X_K = \{x_i \in S \mid i=1, 2, \dots, K\}$ 中各元素的平均 K 近邻距 $nnad(x_i)$ 的计算公式为

$$nnad(x^*) = \frac{\sum_{i=1}^K d(x^*, x_i)}{K} \quad (1)$$

其中

$$d(x^*, x_i) = \sqrt{\sum_{p=1}^n (x_p^* - x_{ip})^2} \quad (2)$$

式中, $d(x^*, x_i)$ 为欧氏距离; x_p^* 、 x_{ip} 分别为点 x^* 与 x_i 的第 p 个测井参数对应的数值。

再次,根据参考样本集选定一个合理的阈值 λ ,根据核心点在簇中的相对密度对初始簇进行扩展,得到目标簇。

对于核心点为 x_{core} 的初始簇 C ,采用如下方法进行扩展。首先计算 x_{core} 在 C 中的相对密度 $rden(x_{core})$,观察其相对密度是否满足条件 $rden(x_{core}) \geq \lambda$ 。若不满足该条件,则将在 C 之内与 x_{core} 距离最远的点剔除出簇,直到满足条件为止;若满足该条件,则将在 C 之外与 x_{core} 距离最近的点加入 C ,直到不满足条件为止。通过扩展得到的簇即为目标簇。其中 x_{core} 在 C 中的相对密度^[21]的计算公式为

$$rden(x_{core}) = \frac{\min_{x_p \in C} \{nnad(x^*), nnad(x_p)\}}{\max_{x_p \in C} \{nnad(x^*), nnad(x_p)\}} \quad (3)$$

λ 为一个不大于 1 的无量纲常数,其选择须根据样本的数据结构而定。一般方法是首先选择一个接近于 1 的初始值,利用已知从属类别的参考样本

对该初始值进行反复验证,直到达到预期聚类效果时为止。 λ 的取值仅影响簇的范围和数量,取值过小会造成误判,取值过大仅会造成运算量的增大,对判别结果影响不大,因此 λ 的选择应秉承“宁大勿小”的原则。

最后,将聚类完毕的目标簇中数据提取出来,并对剩余数据重复进行上述聚类步骤,直到剩下的点无法纳入任何簇内为止。将这些点分别当作单点构成的簇来看待。

1.2 加权 K 近邻投票法识别待测样本

K 近邻法的决策规则为:假设 x_i^* 为类别未知的待测样本点, $X = \{x_j \mid x_j \in \mathbf{R}, i=1, 2, \dots, p\}$ 为由 p 个类别已知的参考样本组成的数据集,对于待测点 x_i^* ,在 X 中寻找与之距离最近的 K 个点,则 x_i^* 归属于其 K 个近邻的类别中占优势者^[22-23]。

对每个簇的核心点,分别与参考样本集中的元素计算加权距离,寻找其 K 个最近邻居进行投票,将核心点归类于其中占优势的类别,并将各簇元素归类于其核心点所属类别。

在测井解释工作中,由于试油、录井及其他资料的限制,各类别参考样本的容量很难做到完全一致,加之地质结构错综复杂,传统的 K 近邻分类法的计算精度很难得出令人满意的结果。针对上述问题,在 K 近邻法的基础上,引入权重系数对近邻距离的计算值进行加权,从而有效地减小了由样本容量的差异导致的误差,实现了提高识别精度的目的^[23]。

待测样本集 X^* 中某一点 x_m^* 与参考样本 X 中属于第 i 类的样本点 x_n 之间的加权距离为

$$d_{\omega}(x_m^*, x_n) = \omega_i d(x_m^*, x_n) \quad (4)$$

其中

$$\omega_i = \exp\left(-\frac{n_i}{n}\right) \quad (5)$$

式中, ω_i 为权重系数; n_i 为第 i 类中的样本数量; n 为所有类别中样本的总数。

经过加权 K 近邻投票,可以得到每个数据点所对应的流体类别,用“1、2、3、4”分别代表“干层、油层、油水同层、水层”,可以绘制出流体指示识别曲线,并依此识别各层段流体类别。

1.3 算法实现流程

DCBKNN 算法的实现流程如图 2 所示。

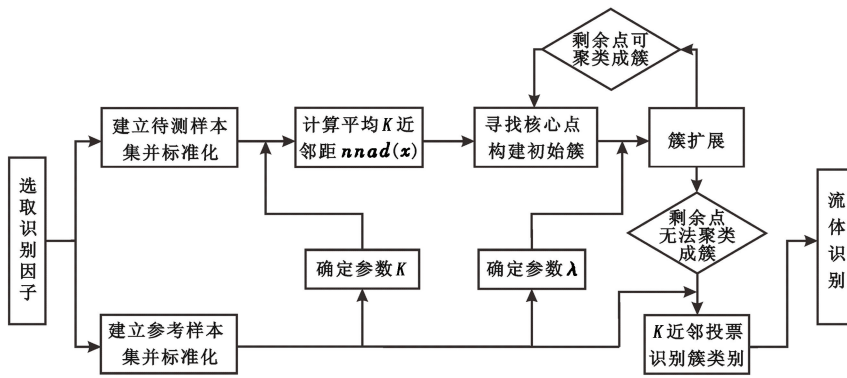


图2 DCBKNN 算法实现步骤流程

Fig.2 Flow-chart of DCBKNN algorithm

2 应用效果分析

选取深侧向电阻率(RD)、浅侧向电阻率(RS)、声波时差(AC)、密度(DEN)、补偿中子(CNL)、自然伽马(GR)等6条常规测井曲线作为识别因子,应用某油田奥陶系鹰山组碳酸盐岩地层24口井共42个试油层段的测试结果,将流体识别类型分为干层、油层、油水同层和水层等4类。

从各层段中提取最能体现该层段流体性质的特征点作为参考样本集。通过对样本集的交叉验证和

反复测试,选定适合本问题的参数 $K=13$ 和 $\lambda=0.92$ 。

采用上述的 DCBKNN 算法对待测样本进行判别。首先将标准化后的 2 209 个待测样本点聚类为 117 个数据簇,其中包括单点构成的簇 79 个。利用加权 K 近邻投票对这 117 个簇的核心点进行识别,得到的结果即为各簇元素的类别。应用结果表明,DCBKNN 算法识别准确度高达 90.48%,如表 1 所示,说明该算法在面对数据量大、结构复杂的储层流体识别问题时,亦能保证较高的识别精度。

表1 DCBKNN 算法识别结果统计

Table 1 Identification result statistics of DCBKNN algorithm

编号	厚度 T/m	GR/ API	RD/ ($\Omega \cdot m$)	RS/ ($\Omega \cdot m$)	AC/ ($\mu s \cdot cm^{-1}$)	DEN/ ($g \cdot cm^{-3}$)	CNL/ %	试油 结论	识别 结果
1	2.0	6.456	749.53	892.00	1.605	2.668	0.595	油层	油层
2	11.0	4.232	521.40	366.44	1.684	2.729	0.866	油层	油层
3	5.0	7.134	778.34	969.42	1.738	2.738	0.817	油层	油层
4	9.0	7.468	428.23	334.74	1.614	2.715	1.362	油层	油层
5	15.0	1.552	544.26	527.45	1.592	2.669	0.723	油层	油层
6	9.0	15.876	372.13	378.93	1.647	2.679	0.757	油层	油层
7	27.0	3.738	943.72	416.84	1.634	2.702	0.713	干层	干层
8	10.0	12.053	865.70	542.71	1.617	2.715	0.835	油层	油层
9	4.0	4.563	899.28	772.18	1.606	2.684	0.316	干层	干层
10	5.0	4.177	503.36	438.01	1.606	2.695	0.526	干层	干层
11	6.0	5.623	379.02	274.91	1.643	2.717	0.892	油层	油层
12	3.0	7.655	315.27	219.06	1.580	2.714	0.496	水层	水层
13	6.0	9.510	625.04	524.61	1.569	2.716	0.374	油水同层	油水同层
14	5.0	10.298	162.77	102.79	1.608	2.714	0.664	水层	水层
15	3.0	7.469	526.71	347.77	1.626	2.671	0.458	水层	水层
16	8.0	3.557	510.78	437.44	1.619	2.681	0.205	油层	油层
17	5.0	29.227	9.26	7.58	2.052	2.616	6.368	水层	水层
18	5.0	7.774	110.00	101.77	1.636	2.693	0.662	水层	水层
19	5.0	11.137	340.59	403.73	1.634	2.709	0.500	油水同层	油水同层
20	5.0	7.120	1 271.20	1 990.60	1.627	2.682	0.274	干层	干层
21	8.0	20.535	54.95	7.82	1.707	2.677	0.751	油层	水层
22	6.5	1.029	596.25	512.64	1.612	2.684	1.023	油层	油层
23	7.0	2.549	295.37	214.51	1.599	2.698	0.203	油层	油层
24	2.5	0.766	997.27	1 103.30	1.606	2.690	0.205	油层	油层
25	7.0	9.476	408.26	275.74	1.686	2.689	0.872	油层	油层
26	4.0	4.637	1 821.20	1 281.00	1.777	2.667	1.045	油水同层	油层

续表

编号	厚度 T/m	GR/ API	RD/ ($\Omega \cdot m$)	RS/ ($\Omega \cdot m$)	AC/ ($\mu s \cdot cm^{-1}$)	DEN/ ($g \cdot cm^{-3}$)	CNL/ %	试油 结论	识别 结果
27	4.0	4.082	331.23	151.55	1.717	2.585	1.073	油层	油层
28	8.0	2.949	450.06	239.56	1.671	2.678	0.660	油水同层	油水同层
29	26.0	4.134	338.62	229.68	1.637	2.679	1.096	油层	油层
30	10.0	10.070	1730.80	2187.80	1.584	2.688	1.971	干层	干层
31	8.0	1.331	525.19	323.19	1.603	2.735	0.064	干层	干层
32	6.0	1.429	664.88	277.79	1.593	2.729	0.099	干层	油层
33	2.5	4.822	715.81	212.99	1.619	2.656	2.047	油层	油水同层
34	5.0	9.651	195.25	186.14	1.668	2.685	0.137	水层	水层
35	6.0	10.021	196.33	165.35	1.656	2.672	0.052	油层	油层
36	10.0	8.867	88.08	56.55	1.689	2.681	0.244	水层	水层
37	2.0	3.838	625.79	443.74	1.614	2.679	0.730	油水同层	油水同层
38	5.0	3.340	326.93	214.49	1.608	2.696	0.723	油层	油层
39	5.0	2.586	134.99	58.83	1.623	2.681	1.534	水层	水层
40	6.0	5.055	326.13	228.98	1.594	2.693	0.452	油层	油层
41	2.2	7.078	126.17	139.42	1.658	2.659	1.942	水层	水层
42	4.5	5.876	702.40	380.34	1.548	2.690	1.164	油层	油层

此外,在识别其他方法难以准确识别的油水同层方面,DCBKNN 算法仍具有一定的优势。在参与测试的 42 个层段中,有油水同层段 5 个(层段 13、19、26、28、37),其中仅有 1 个层段判断错误(层段 26),识别

正确率达 80%。以同一口井中的层段 28、29 为例进行说明,见图 2。从测井响应来看,层段 28(图 3 中层段 1)平均深侧向电阻率为 450.06 $\Omega \cdot m$,具有小幅正差异,声波、中子曲线小幅升高,密度曲线小幅降低,

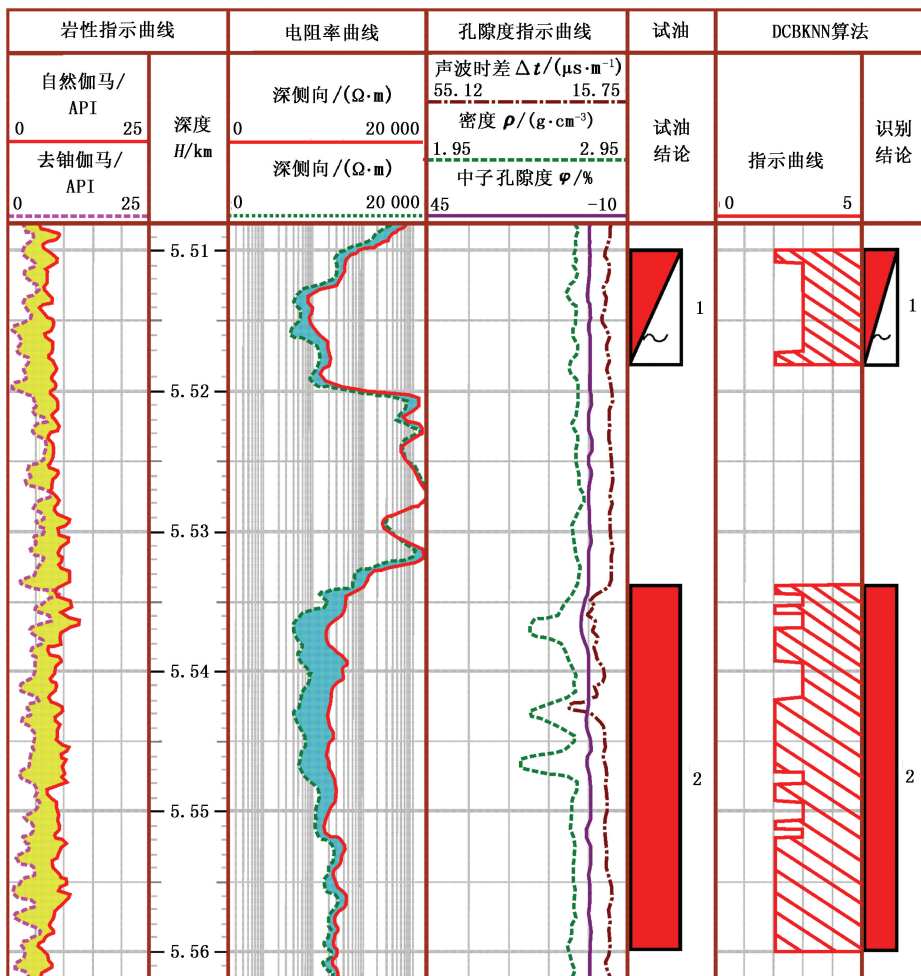


图 3 DCBKNN 在层段 28、29 的识别结果

Fig.3 Identification results of DCBKNN on interval 28 and 29

整体表现为油水同层特征;层段 29(即图 3 中层段 2)平均深测向电阻率为 $328.62 \Omega \cdot \text{m}$,具有较大幅度正差异,声波、中子曲线升高幅度和密度曲线降低幅度均十分明显,整体表现为油层。DCBKNN 算法的结果用“1、2、3、4”分别代表“干层、油层、油水同层、水层”,通过运算得出的识别曲线,判断层段 28、29 分别为油水同层和油层,与试油结论完全吻合。由此可见,DCBKNN 算法在识别油水同层等复杂储层时具有较好的效果。

3 结 论

(1)基于密度聚类的 K 近邻法是对 K 近邻法和密度聚类算法的改进,引入权重函数以减小样本比重对 K 近邻法的影响,引入相对密度以弥补密度聚类算法参数敏感性过高的缺点。利用相对密度聚类对 K 近邻法的数据加以简化,提高了运算效率,在处理数据量大、结构复杂的问题时,亦能保证较高的识别精度。

(2)基于密度聚类的 K 近邻法是一种非参数的识别方法,既能有效解决传统识别方法过于依赖个人主观经验的问题,又能克服其他常用的数据挖掘和模式识别方法精度低、运算慢、易产生过拟合的缺点,提高了算法的泛化性和鲁棒性。

(3)基于密度聚类的 K 近邻法能更好地识别其他常用算法难以识别的油水同层、油气同层等复杂层段,具有良好的应用前景,为利用数据挖掘方法解决油田勘探开发中的复杂问题提供了新思路。

参考文献:

[1] 赵克超,陶果,王天波,等. 泌阳凹陷深层系低孔低渗油气层测井识别技术[J]. 石油大学学报:自然科学版,2005,29(6):27-31.
ZHAO Kechao, TAO Guo, WANG Tianbo, et al. Hydrocarbon identification of the deep-seated reservoirs with low porosity and low permeability from well logs in Biyang sag [J]. Journal of the University of Petroleum, China (Edition of Natural Science), 2005, 29(6): 27-31.

[2] 司马立强,郑淑芬,罗宁,等. 川东地区石炭系储层流体性质测井判别方法适应性分析[J]. 天然气工业, 2012, 22(6): 45-48.
SIMA Liqiang, ZHENG Shufen, LUO Ning, et al. Adaptability analysis of discriminating the property of fluids in carboniferous reservoirs in east Sichuan by logging methods [J]. Natural Gas Industry, 2012, 22(6): 45-48.

[3] 徐德龙,李涛,黄宝华,等. 利用交会图法识别国外 M

油田岩性与流体类型的研究[J]. 地球物理学进展, 2012, 27(3): 1123-1132.

XU Delong, LI Tao, HUANG Baohua, et al. Research on the identification of lithology and fluid type of foreign oilfield by using the crossplot method [J]. Progress in Geophys, 2012, 27(3): 1123-1132.

[4] 邹文,贺振华,陈爱萍,等. 定量交会图技术及其在流体识别中的应用[J]. 石油物探, 2008, 47(1): 45-48.
ZOU Wen, HE Zhenhua, CHEN Aiping, et al. Application of quantitative crossplot technique in fluid identification [J]. Geophysical Prospecting for Petroleum, 2008, 47(1): 45-48.

[5] 陈刚. 一种新的识别低电阻率油层的交会图及其应用 [J]. 中国石油大学学报:自然科学版, 2008, 32(3): 36-39.
CHEN Gang. A new cross plot for identifying low resistivity reservoirs and its application [J]. Journal of China University of Petroleum (Edition of Natural Science), 2008, 32(3): 36-39.

[6] 陈钢花,张蕾,宋国奇,等. 测井资料在地层不整合纵向结构研究中的应用[J]. 中国石油大学学报:自然科学版, 2010, 34(1): 50-54.
CHEN Ganghua, ZHANG Lei, SONG Guoqi, et al. Application of logging data to unconformable vertical structure research of stratum [J]. Journal of China University of Petroleum (Edition of Natural Science), 2010, 34(1): 50-54.

[7] 袁宇,高楚桥,邹雯,等. 基于交会图和全概率公式的流体识别方法[J]. 石油天然气学报, 2013, 35(5): 79-82
YUAN Yu, GAO Chuqiao, ZOU Wen, et al. Method of fluid identification based on crossplot and total probability formula [J]. Journal of Oil and Gas Technology, 2013, 35(5): 79-82.

[8] 李梅,赖强,黄科,等. 低孔低渗碎屑岩储层流体性质测井识别技术:以四川盆地安岳气田须家河组气藏为例[J]. 天然气工业, 2013, 33(6): 34-38.
LI Mei, LAI Qiang, HUANG Ke, et al. Logging identification of fluid properties in low porosity and low permeability clastic reservoirs: a cases study of Xujiahe Fm gas reservoirs in the Anyue Gas Field, Sichuan Basin [J]. Natural Gas Industry, 2013, 33(6): 34-38.

[9] 樊家琨. 应用多元分析[M]. 开封:河南大学出版社, 1993: 135-144.

[10] 罗德江. 基于核 Fisher 判别的碎屑岩储层流体识别 [J]. 地球物理学进展, 2013, 28(4): 1919-1924.
LUO Dejiang. The method on the fluid identification of rock elastic properties by Fisher discriminant analysis in

- fragmental rock[J]. *Progress in Geophys*,2013,28(4):1919-1924.
- [11] 黄烈林,侯健,陈月明,等. Fisher 判别法在聚合物驱替潜力评价中的应用[J]. *石油大学学报:自然科学版*,2002,26(1):49-51.
HUANG Lielin, HOU Jian, CHEN Yueming, et al. Application of Fisher discrimination method to potential evaluation of polymer flooding[J]. *Journal of the University of Petroleum, China(Edition of Natural Science)*, 2002,26(1):49-51.
- [12] 刘立峰,孙赞东,韩剑发,等. 量子粒子群模糊神经网络碳酸盐岩流体识别方法研究[J]. *地球物理学报*,2014,57(3):991-1000.
LIU Lifeng, SUN Zandong, HAN Jianfa, et al. A carbonate fluid identification method based on quantum particle swarm fuzzy neural network[J]. *Chinese Journal of Geophysics*,2014,57(3):991-1000.
- [13] 岳文正,陶果. 地球物理测井多参数综合识别储层流体类型的新型神经网络[J]. *石油大学学报:自然科学版*,2004,28(3):30-42.
YUE Wenzheng, TAO Guo. A new neural network applied to comprehensive recognition of reservoirs fluid type using geophysical well logging parameters[J]. *Journal of the University of Petroleum, China(Edition of Natural Science)*,2004,28(3):30-42.
- [14] 李学慧. 苏北盆地低电阻率油层的识别及其分布[J]. *中国石油大学学报:自然科学版*,2008,32(5):1-6.
LI Xuehui. Identification of the low resistivity reservoir and its distribution in Subei Basin[J]. *Journal of China University of Petroleum(Edition of Natural Science)*, 2008,32(5):1-6.
- [15] 孟庆民,张士诚,王莉,等. 数据挖掘技术在气田压裂效果评价中的应用[J]. *中国石油大学学报:自然科学版*,2008,32(5):165-169.
MENG Qingmin, ZHANG Shicheng, WANG Li, et al. Application of data mining to postfracture response evaluation in gas field[J]. *Journal of China University of Petroleum(Edition of Natural Science)*, 2008,32(5):165-169.
- [16] 潘和平,樊政军,马勇. 基于信息熵识别油气层和水层的聚类方法[J]. *石油大学学报:自然科学版*,2004,28(6):31-34.
PAN Heping, FAN Zhengjun, MA Yong. Clustering method for identification of oil(gas)-bearing formation and water-bearing formation based on information entropy[J]. *Journal of the University of Petroleum, China(Edition of Natural Science)*, 2004,28(6):31-34.
- [17] 印兴耀,叶端南,张广智. 基于核空间的模糊聚类方法在储层预测中的应用[J]. *中国石油大学学报:自然科学版*,2012,36(1):53-59.
YIN Xingyao, YE Duannan, ZHANG Guangzhi. Application of kernel fuzzy C-means method to reservoir prediction[J]. *Journal of China University of Petroleum(Edition of Natural Science)*, 2012,36(1):53-59.
- [18] 赵杰,杨柳. 聚类分析的 dBscan 的改进与实现[J]. *微电子学与计算机*,2009,26(11):189-192.
ZHAO Jie, YANG Liu. The improvement and implementation of dBscan clustering algorithm[J]. *Microelectronics & Computer*,2009,26(11):189-192.
- [19] 冯少荣,肖文俊. DBSCAN 聚类算法的研究与改进[J]. *中国矿业大学学报*,2008,37(1):105-111.
FENG Shaorong, XIAO Wenjun. An improved DBSCAN clustering algorithm[J]. *Journal of China University of Mining & Technology*,2008,37(1):105-111.
- [20] 李航. 统计学习方法[M]. 北京:清华大学出版社,2012:37-44.
- [21] 孙凌燕,杨明,任建斌. 一种基于相对密度的快速聚类算法[J]. *微电子学与计算机*,2009,26(12):109-116.
SUN Lingyan, YANG Ming, REN Jianbin. A fast density-based clustering algorithm[J]. *Microelectronics & Computer*,2009,26(12):109-116.
- [22] 王淑盛,徐正光,刘黄伟,等. 改进的 K 近邻方法在岩性识别中的应用[J]. *地球物理学进展*,2004,19(2):478-480.
WANG Shusheng, XU Zhengguang, LIU Huangwei, et al. The advanced k-nearest neighborhood method used in the recognition of lithology[J]. *Progress in Geophysics*,2004,19(2):478-480.
- [23] 曹根,葛孝堃,杨丽琴. 基于 K-近邻法的局部加权朴素贝叶斯分类算法[J]. *计算机软件与应用*,2011,28(9):267-291.
CAO Gen, GE Xiaokun, YANG Liqin. Locally weighted naive Bayes classification algorithm based on K-nearest neighborhood[J]. *Computer Applications and Software*, 2004,19(2):478-480.

(编辑 修荣荣)