

基于聚类分析的石油储层中含油性识别方法

张仲阳, 高胜利

西安石油大学, 地球科学与工程学院, 陕西 西安

收稿日期: 2023年10月18日; 录用日期: 2023年12月21日; 发布日期: 2023年12月28日

摘要

储层含油性预测是油藏描述的重要内容之一。用聚类分析方法对测井资料进行油气、水层识别的思路。油储层的预测即是在井位处获取测井数据, 进行储层含油性识别, 从而得到储层在一个井位的精确认识, 可以使得识别更加客观, 进而得到对储层的完整认识, 而选择一个合适的算法进行属性提取, 可以提高方法的准确性和高效性。所以本文将重点研究, 基于K-Means聚类分析的石油储层中含油性识别。

关键词

软计算, K-Means, 油藏含油性识别, 聚类

Identification Method of Oil Content in Petroleum Reservoir Based on Cluster Analysis

Zhongyang Zhang, Shengli Gao

School of Geosciences and Engineering, Xi'an Shiyou University, Xi'an Shaanxi

Received: Oct. 18th, 2023; accepted: Dec. 21st, 2023; published: Dec. 28th, 2023

Abstract

Reservoir oiliness prediction is one of the important contents of reservoir description. The method of cluster analysis is used to identify oil, gas and water layers from well logging data. The prediction of oil reservoir is to obtain logging data at the well location and identify the oil content of the reservoir, so as to obtain an accurate understanding of the reservoir at a well location, which can make the identification more objective and obtain a complete understanding of the reservoir. In addition, selecting a suitable algorithm for attribute extraction can improve the accuracy and efficiency of the method. Therefore, this paper will focus on the identification of oil content in petroleum reservoirs based on K-Means cluster analysis.

Keywords

Soft Computing, K-Means, Oil Reservoir Identification, Clustering

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

对石油行业的数据分析揭示了中国对外部石油的依赖程度每年都呈上升趋势。尽管我国石油产量在逐年递增,但是国内经济的快速增长导致了石油消费的急速上升,这使得石油产量的年度增长量相对较小,从而导致对外部石油的依赖度持续增长,进一步扩大了我国石油供应和需求之间的不平衡[1]。因此,为了缓解石油供需的压力,增大储备和生产无疑是必要的步骤。在这种背景下,复杂的石油储层特性,如非均匀、非线性和不确定性反应特性,以及石油储层管理中的大量测井数据,其信息量大,来源广泛。在这种情况下,对已有的测井数据的深入分析和应用,准确识别石油储层变得非常重要[1]。

将基于聚类分析算法应用于石油储层识别中,能够得到储层识别中的关键测井属性,用这些关键属性进行石油储层的分类有助于提高石油储层的识别率,同时能够约减一些次要的测井属性,减少在石油勘探阶段需要进行勘探的属性数量以便节约开采成本。

K-Means 聚类算法在地球科学、信息技术、决策科学等众多领域有着广泛的应用前景。K-Means 聚类的核心是将多个维度数据集按一定的距离划分成 K 个类别,这样在同一类别内的对象之间存在较大的相似度,而在不同类内的对象之间存在较大的差别[2]。

如何实现油气层特征的快速、精确辨识,是目前国际上石油资源勘查与开发中急需攻克的难点问题。为了解决这个问题,发展出了以测录井数据为主要手段的方法,然而,对于复杂岩性和低渗透率的储层,其适用范围还不够广。K-Means 聚类的发展为油气层识别工作提供了新的方法和途径。本文从实际储层出发,采用 K-Means 聚类的思想,将测井资料和岩芯测试结果有机地融合起来,在选取特征参数的基础上,采用 K-Means 聚类分析方法,对陕北吴起白河区块的油气藏开展研究。

2. K-Means 聚类分析模型

2.1. K-Means 算法原理

聚类是一种依据各类特性将数据分拆至不同的集群或群组的步骤,属于同一群组的元素具有相似性,然而,不同群组之间的元素相差甚远。简单来说,聚类分析是在缺乏预知的状况下,利用各种方法根据数据自身的类似性将数据点分配至不同的集群的过程[3]。

K-Means 算法,也被称为 K-均值算法,是在无监督学习域里广泛应用的一种聚类算法。此方法采用物体之间的相似度作为分类数据的依据,通过计算距离函数,比较样本 P 和样本 Q 之间的相似性,以此将极具相似性的数据分类至同一个集群。其关键思路在于:在给定 K 值以及 K 个初始类群中心的基础上,定义每个数据记录(即点)属于距离自身最近的类群中心所代表的类别。当所有的数据点都已被归属后,根据每个类群中的所有数据点,重新计算该类群的中心点(求平均值);然后再继续处理这些分配点和更新类群中心的步骤,直到类群中心的改变范围微乎其微,或者达到预设的迭代次数为止[4][5]。

我们假设有一个数据样本 X , 它由 n 个元素组成, 即 $X = \{X_1, X_2, X_3, \dots, X_n\}$, 每个元素都具备 m 个维度的特性。K-Means 算法的目的在于, 将这 n 个元素基于他们间的相似性分配到预设的 K 个簇中, 每个元素只能被分配到离它最近的簇。在应用 K-Means 算法之前, 首先要确定 k 个初始聚类中心 $\{C_1, C_2, C_3, \dots, C_k\}$, $1 < k \leq n$, 然后通过计算每一个对象到每一个类中心的欧氏距离, 如下式所示 $dis\{X_i, C_j\} = \sqrt{\sum_{t=1}^m (X_{it} - C_{jt})^2}$, X_i 表示的是第 i 个对象 $1 \leq i \leq n$, C_j 表示的是第 j 个集群中心, j 的范围是 $1 \leq j \leq k$, X_{it} 指的是第 i 项的第 t 个特性, t 的范围是 $1 \leq t \leq m$, C_{jt} 则表示的是第 j 个集群中心的第 t 个特性。根据此方法, 比较每个项目与每个集群中心的距离, 将项目分类到与其最近的集群之中, 从而形成 k 个不同的分组, 该方法就被称为 K-Means 算法[6] [7]。

K-means 算法的初始聚类中心是随机选取的, 而聚类结果易受初始中心的影响, 因此, 采用不同的初始聚类中心进行聚类, 所得的聚类效果也存在较大差异。在此基础上, 本文先求出数据集中的每一个数据点到原点的距离, 再将这些数据点按照此距离进行排序, 把已排列好的数据平均分成 k 个组, 并以各组的中间数据点为初始聚类中心。由此, 可以减少聚类结果误差过大影响其效果。

2.2. K-Means 算法流程

首先, 根据需要划分的簇类数 k , 从样本集合中选取 k 个数据中心, 对每个样本进行多次迭代, 并计算每一个样本对 k 个数据中心的距离。当一次遍历结束时, 全部的样本都找到自身属于的类别, 但是这个分类并不是最后的类别, 因此需要再做一次循环。在第二遍迭代前, 首先要在簇中更新质心, 计算出每一个簇中的质心, 再反复上述过程, 直至到达期望的迭代数目或多次迭代的结果, 簇内的信息都不会改变[8]。K-Means 均值算法的主要过程如图 1 所示。

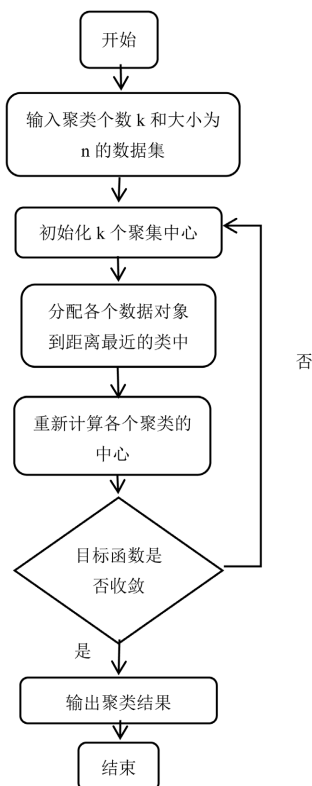


Figure 1. K-Means basic flow of the algorithm
图 1. K-Means 算法基本流程

K-Means 聚类算法是一个不断迭代的过程, 原始数据集有 4 个簇, 图中 x 和 y 分别代表数据点的横纵坐标值, 采用 K-Means 方法对样本进行聚类, 并将其进行两次迭代, 得出最后的聚类结果, 迭代过程如图 2 所示。

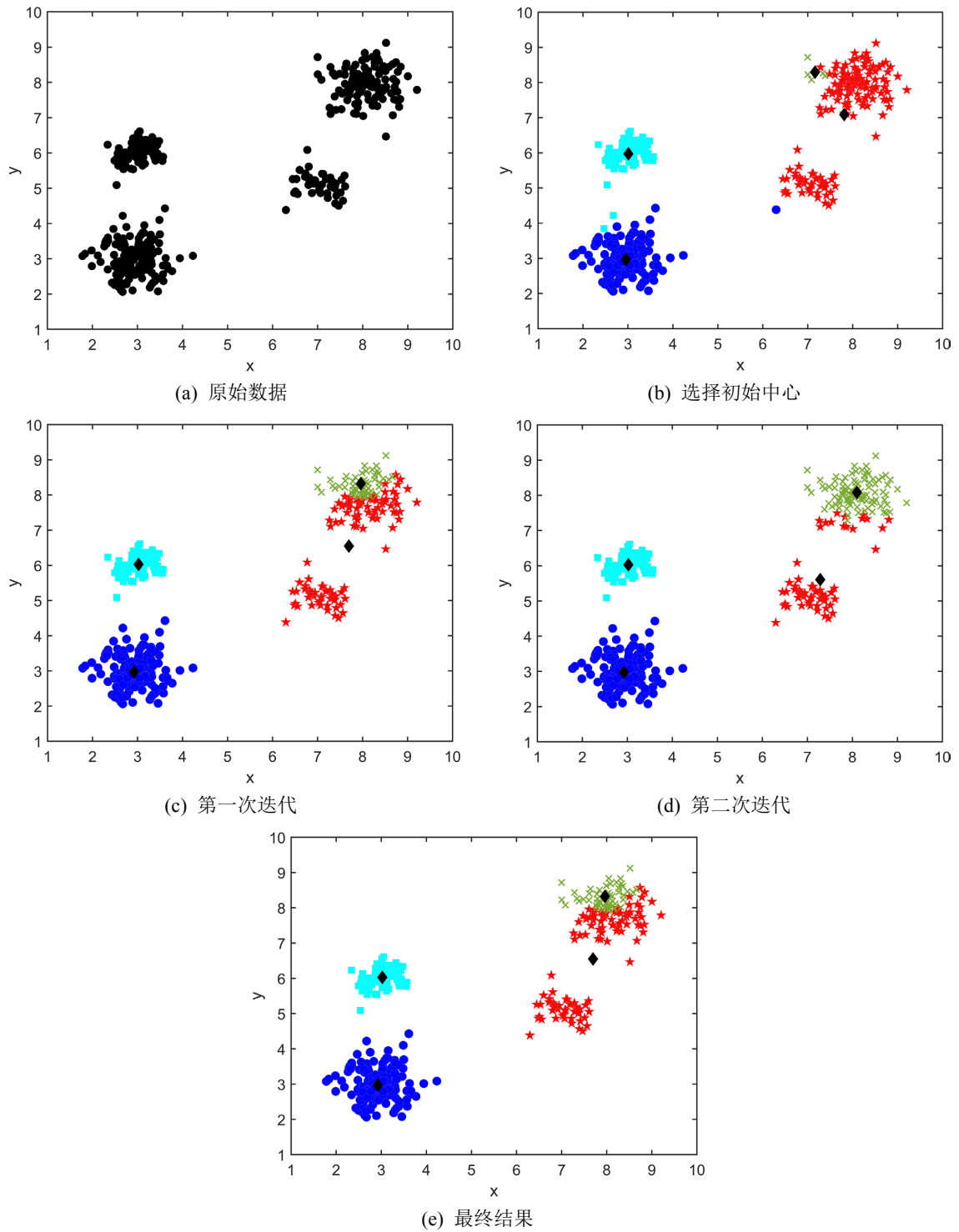


Figure 2. K-Means algorithm iteration process

图 2. K-Means 算法迭代过程

3. 储层含油性识别中的 K-Means 聚类

3.1. 样本和指标的选取

本次样本和指标所选取的数据源自于陕北吴起白河区域。吴起白河油区位于吴起县城西北部, 东西长约 15 km, 南北宽约 11 km, 面积约 110 km²。研究区目的层构造活动较弱, 构造形态简单, 处于一级构造单元伊陕斜坡带的西部(图 3)。在此背景上发育了多排局部略高的微幅度鼻状隆起。这些微幅度鼻状隆起, 在平面上呈隆凹相间的条带状分布。地层岩性多呈砂泥岩互层, 受到成岩期压实作用影响, 纵向上各层构造具继承性发育的特点, 但横向上各层构造高点略有偏移[9][10]。

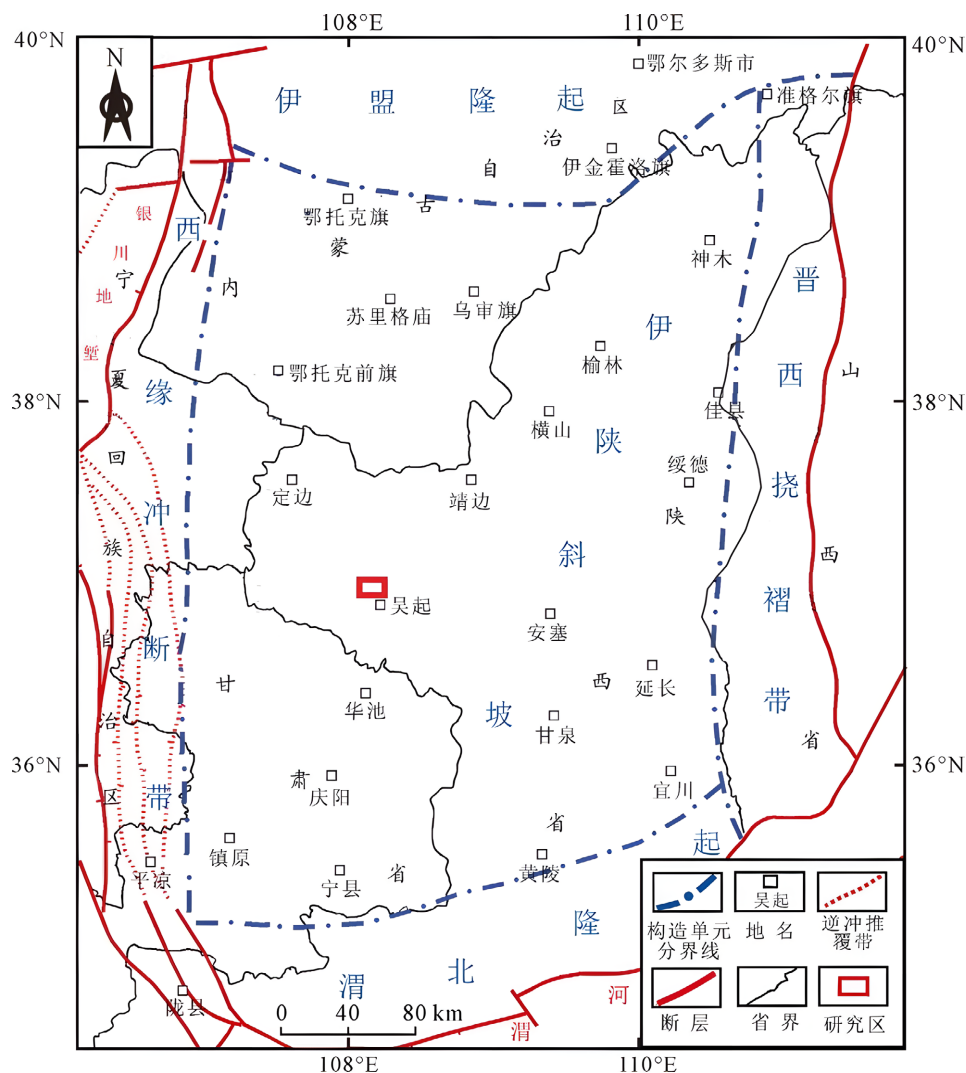


Figure 3. Regional map of structural location in the study area (Xiong Anliang 2023)

图 3. 研究区构造位置区域图(引自熊安亮 2023)

根据沉积旋回与储层非均质性的关系, 因此在进行地层划分时, 需要采用测井资料进行地层划分。测井资料的数据量越大, 对储层的物性、流体特性的刻画也就越来越细致, 所以用测井曲线来进行地层划分是最基本、最有效的方法。通过对鄂尔多斯盆地延长组地层(表 1)的分析, 发现长 3 号、长 4+5 和长 6 的测井资料具有“高电位, 低密度, 低电阻”的特征[11][12]。

Table 1. Marker surface layer in Yanchang formation
表 1. 延长组地层中的标志层

地层					厚度(m)	岩性特征	电性特征
系	组	段	油层组	油层亚组			
三 叠 系	延 长 组		长 3		70~110	灰绿色泥岩与灰绿色砂岩	高电阻低声波
		T ₃ yc ₄	长 4 + 5	1	40~45	含油细砂岩夹薄层泥岩	高电阻低声波
				2	40~45	薄层状细砂岩与灰黑色泥岩互层	高电阻低声波
		T ₃ yc ₃	长 6		100~110	中细砂岩夹灰黑色泥岩	上部低电阻高声波、 中下部低电阻低声波

实验数据来源于陕北吴起白河区域, 目前储层和微构造条件是石油聚集的关键因素, 该井的测井特性和相关的测井解读成果可在表 2 中查询。

Table 2. Logging interpretation and identification results
表 2. 测井解释结论和识别结果

ID	起始深度	结束深度	统计项	自然电位 SP	双感应(深)RILD	孔隙度 / (%)	渗透率 / μm^2	测井结论	聚类结果
长 3	1964.8	1978.3	平均值	39.2	5.9	9.5	2.1	水层	水层
			最大值	42.2	6.7	11.5	3.9		
			最小值	35.8	5.1	0	0		
长 3	1990.7	1993	平均值	37.3	6.4	4.8	1.6	水层	水层
			最大值	40.8	7	12.1	4.9		
			最小值	35.7	5.9	0	0		
长 3	1993	1994.6	平均值	39.8	6.3	4.4	1.2	干层	干层
			最大值	41.5	6.8	10.7	2.9		
			最小值	36.3	5.9	0	0		
长 3	1994.6	2008.8	平均值	28.2	5.3	9.3	1.8	水层	水层
			最大值	35.8	8.7	11.9	4.5		
			最小值	22.6	4.1	0	0		
长 3	2018.6	2033.2	平均值	27	6.7	8.4	1.9	水层	水层
			最大值	34.4	10.8	13.3	7.5		
			最小值	18	4.1	0	0		
长 4 + 5	2057.9	2060.7	平均值	44	11.3	1.9	0.3	水层	水层
			最大值	51.8	13	9.4	1.6		
			最小值	40.2	10.1	0	0		

Continued

			平均值	21.8	6.3	12.2	5.1		
长 4+ 5	2075.7	2077.7	最大值	24.3	6.4	12.6	5.8	干层	干层
			最小值	20.3	6.2	11.9	4.6		
			平均值	61.7	12	1	0.3		
长 4+ 5	2112.7	2114.2	最大值	62.7	12.6	11.4	3.9	水层	水层
			最小值	59.2	11.5	0	0		
			平均值	40	5.9	15.4	20.2		
长 4+ 5	2115.4	2116.9	最大值	46.2	7.9	21.5	61.7	水层	水层
			最小值	36.7	5.1	12.1	4.9		
			平均值	73.8	14.4	0.5	0.1		
长 6	2166.9	2169.1	最大值	74.6	16.4	8.4	1	致密油层	致密油层
			最小值	73.2	12.5	0	0		
			平均值	70.6	15.8	0	0		
长 6	2171.2	2171.9	最大值	71.7	17.3	0	0	致密油层	致密油层
			最小值	69.4	14.6	0	0		
			平均值	67.2	18.5	4.7	0.4		
长 6	2172.4	2173.7	最大值	67.7	20	8.3	0.9	致密油层	致密油层
			最小值	65.9	17.3	0	0		
			平均值	63.1	20.1	7.1	1.4		
长 6	2173.7	2174.4	最大值	65.2	23.7	11.5	4	干层	水层
			最小值	61.3	17.7	0	0		
			平均值	45.3	12.5	7.9	1.4		
长 6	2237.1	2250.5	最大值	57.8	18.8	11.7	4.3	干层	干层
			最小值	27.6	7.4	0	0		

我们可以从表 2 的数据中了解到, 用于识别储层含油性的几种特征包括: 自然电位、孔隙率、双感应以及渗透率。井测得出的结论指出: 致密油层、水层和干层。在进行聚类的过程中, 利用欧氏距离来评价相似性, 并进行了归一化处理以消除无量纲下数值差异。所有的 4 个属性, 都被用于储层含油性的辨识。

3.2. 分类结果

表 2 的最后一栏展示了聚类的结果, 其中用粗体标注了与测井结果有出入的样例。经过比对发现, 鉴别的误差大致集中在差油层和水层以及油层和差油层的交界处[13]。这主要源于界定差油层和油层、差

油层和水层的度量标准[14]。实际上, 由于很难确定精确的分界标准, 在实际操作中通常只能给出一个可以接受的范围。但是即便如此, 整体上聚类效果有较高的准确率, 对于预测测井数据仍具有很高的价值。

4. 结论

(1) 储层含油性的确定是油气藏测井评估中的一个关键环节, 是把测井数据转化为地质信息, 反映测井解释结果及使用效果的更高阶段, 其好坏将直接关系到油气勘探的效果和成功率。在某些复杂的低孔渗、低电阻率等地质条件下, 常规的测井解释方法难以取得理想的效果。本文提出将 K-Means 聚类分析的方法引入到储层识别中, 使其可以较好地识别储层的各种复杂特征, 符合应用要求[15][16]。

(2) 油气藏的辨识是一个多因素的综合分析, 单独利用一种数据往往无法达到预期的效果, 尤其是对于具有复杂性质的油气藏, 其辨识成果很难为油气藏的勘探与开采提供依据。因此, 在实践中如何选取合适的典型参量还有待于进一步提炼和归纳。K-Means 聚类方法用于储层含油性的识别时, 由于储层类型数量的变化而有所不同, 聚类质量差异很大, 因此 K-Means 聚类方法在实际中具有很大的局限性, 容易陷入局部极值。

(3) 自 K-Means 算法提出以来至今已有多时, 它已经成为一种非常典型的聚类方法, 并且在未来的研究中会得到更广泛的应用。对于传统 K-Means 算法是聚类分析中一种常用的基于划分的方法, 其不足之处在于其对初始聚类过于敏感。为此, 本文针对这一点进行了优化, 将初始聚类中心的随机选取, 优化为对各数据点到原点的距离进行排序后, 选择 k 个数据点作为初始聚类中心, 可减少聚类误差大影响其效果。

(4) K-Means 方法在油气藏辨识领域具有广阔的发展空间, 但仍存在许多问题有待解决。在本项目提出的基础上, 还需要进一步提高在大规模或者高维度数据集条件下的处理能力并且减少其算法的时间复杂度。在此基础上, 我们还需要对其进行更深入的研究, 以提高其识别性能。

致 谢

感谢高胜利教授提供的研究资料以及指导建议。

参考文献

- [1] 李亚楠. 石油储层含油性识别的特征选择和规则提取[D]: [博士学位论文]. 武汉: 中国地质大学, 2018.
- [2] 赵玄, 严家斌, 胡涛. 聚类分析在地球物理中的应用进展[J]. 中国科技信息, 2018(15): 103-105.
- [3] 苏晓眉, 张涛, 李玉飞, 等. 基于 K-Means 聚类算法的沉砂卡钻预测方法研究[J]. 钻采工艺, 2021, 44(3): 5-9.
- [4] 司福明. 一种基于密度的增量 K-Means 聚类算法研究[J]. 长春工程学院学报(自然科学版), 2016, 17(2): 99-102.
- [5] 张琳, 陈燕, 汲业, 等. 一种基于密度的 K-Means 算法研究[J]. 计算机应用研究, 2011, 28(11): 71-73.
- [6] 杨俊闯, 赵超. K-Means 聚类算法研究综述[J]. 计算机工程与应用, 2019, 55(23): 7-14+63.
- [7] 周爱武, 陈宝楼, 王琰. K-Means 算法的研究与改进[J]. 计算机技术与发展, 2012, 22(10): 101-104.
- [8] 黄震华, 向阳, 张波, 王栋, 刘啸岭. 一种进行 K-Means 聚类的有效方法[J]. 模式识别与人工智能, 2010, 23(4): 517-521.
- [9] 蔡绪森, 王胜, 高雯, 于向前, 汪义莉. 吴起油田旗 56-122 区剩余油分布研究[J]. 地球科学前沿, 2022, 12(6): 867-887. <https://doi.org/10.12677/AG.2022.126084>
- [10] 汪义莉. 吴起油田 G 区块长 91 油层组储层特征研究及储层评价[J]. 地球科学前沿, 2022, 12(1): 9-14. <https://doi.org/10.12677/AG.2022.121002>
- [11] 王伟龙, 朱锦艳. 吴起油田 CJN 油区长 4 + 5 油藏分布特征[J]. 云南化工, 2021, 48(5): 138-140.
- [12] 韦悦, 郭峰, 彭晓霞, 等. 吴起油田西部长 4 + 5 致密砂岩储层敏感性特征[J]. 中外能源, 2022, 27(4): 36-44.
- [13] 冯鑫. 吴起油田油沟油区长 4 + 5 油藏精细描述[D]: [硕士学位论文]. 西安: 西安石油大学, 2017.

- [14] 郑卓, 王建立, 夏良冰, 等. 基于高斯混合模型聚类分析的解释方法在埕北油田气顶区混合油气层的应用[J]. 录井工程, 2022, 33(3): 73-77+83.
- [15] 袁可红, 李艳晓, 郭海湘, 等. 一种用于储层含油性识别的蚁群聚类算法[J]. 计算机工程, 2011, 37(13): 178-180.
- [16] 司福明. 猫群与 K-Means 算法在储层含油性识别问题中比较分析[J]. 喀什大学学报, 2016, 37(3): 50-53.