

基于 KNN 的剩余油形态识别

程小龙,王正勇,滕奇志

(四川大学 电子信息学院 图像信息研究所,四川 成都 610065)

摘要:对从实验中采集到的剩余油图像进行分析研究,可以为油藏后期开采提供理论依据。通过收集确定类型的剩余油特征数据作为样本集向量空间,对待分类剩余油特征数据进行归一化处理,之后求取欧氏距离。使用 KNN(K 近邻)分类方法近邻投票确定剩余油类别,可以较为快速准确地得到分类结果。

关键词:剩余油分类;KNN;形态识别

中图分类号:TP312

文献标识码:A

DOI: 10.19358/j.issn.2096-5133.2020.01.020

引用格式:程小龙,王正勇,滕奇志.基于 KNN 的剩余油形态识别[J].信息技术与网络安全,2020,39(1):104-107.

K-nearest neighbor method for recognizing the shape of residual oil

Cheng Xiaolong, Wang Zhengyong, Teng Qizhi

(Institute of Image Information, School of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract:The analysis of the remaining oil images collected from the experiment can provide a theoretical basis for the later stage of reservoir development. In this paper, the remaining oil characteristic data of the determined type is collected as the sample set vector space, and the Euclidean distance is obtained after normalizing the residual oil characteristic data to be classified. According to the KNN (K-nearest neighbor) classification method, the neighbor voting determines the category, and the classification result can be obtained quickly and accurately.

Key words: remaining oil classification; K-nearest neighbor; morphological recognition

0 引言

石油地质研究人员一般通过微观驱替实验来研究采收效率,该实验是对玻璃刻蚀模型进行驱替仿真并在驱替过程中收集驱替图像,实验过程如图 1 所示^[1-2]。剩余油顾名思义是剩余的油,当驱替结束时,孔隙空间中尚残余的油即为剩余油。研究剩余油的形态分类,可对充分挖掘油藏潜力,提高油田采收效率提供理论支撑,为此本文提出了基于 KNN 的分类方法对剩余油形态进行分类^[3-5]。

和膜状,命名虽然不同,但是所分的类别可能相同^[6-8]。本文将剩余油分为孤岛状、凝块状、膜状、网状以及条带状,其形状如图 2 所示。

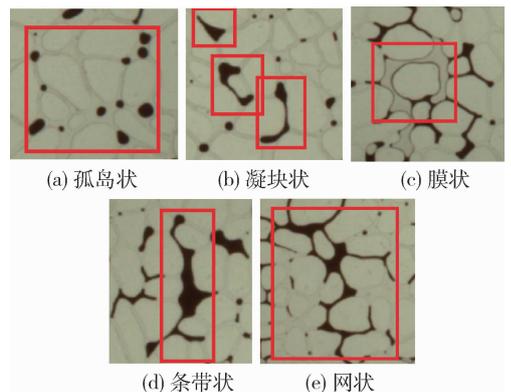


图 2 剩余油类别



图 1 水驱油实验过程

1 剩余油形态类别介绍

剩余油在孔隙中的分布是多种多样的,目前石油地质行业对剩余油的分类类别也没有标准定论,还处于摸索阶段。已经存在的分类类别有单孔型、油膜型、多孔型、连片型,或者是簇状、盲状、孤岛状

孤岛状一般孤立出现在孔隙交汇点,呈类圆形,面积比较小。凝块状出现在孔隙交汇点并且会占据孔隙的一部分,根据面积大小,占据的孔隙空间有多有少,形状呈柱状、椭圆状、塞状等。膜状是

油附在孔隙两边,因为部分岩石亲油而难以用驱替液驱出,所以呈现薄膜状。条带状顾名思义是油呈狭长条带状,形成原因是其两端喉道很窄,驱替液难以在其两端施加压力。网状也即油呈现网络状,分支较多,也是由于喉道压力小,难以驱动所形成。

2 剩余油特征参数介绍

常见的几何特征数据有圆形度、凹度、偏心率、长宽比、周长、面积、长径、短径、最大 Feret 直径、最小 Feret 直径、矩形度、凸周长、Feret 比、径比等。通过观察对比每类剩余油数据,发现圆形度、凹度、长宽比、最小 Feret 直径、周长、面积、Feret 直径比这七种数据对于区分剩余油有比较好的作用。因此,收集待分类剩余油的这七项数据,再用 KNN 算法实现分类。以下是对这些参数的介绍。

(1) 圆形度

圆形度是表示目标区域与圆的近似程度,其公式如式(1)所示, C 表示圆形度, A 表示目标区域面积, P 表示目标区域的周长。当 C 为1时,目标区域为标准圆形;越接近于1,与圆的相似程度越高;越小,与圆的相似程度越低,其形状越复杂。因为孤岛状近似于圆形,所以孤岛状的圆形度很高,圆形度对孤岛状有很好的区分作用。

$$C = \frac{4\pi A}{P^2} \quad (1)$$

(2) 凹度

凹度反映了目标边界的凹凸性,设边界的凸周长为 P_c ,则凹度计算如式(2)所示,式中 P 表示目标边界的周长。

$$C_o = \frac{P}{P_c} \quad (2)$$

(3) 长宽比

长宽比是目标区域的最小外接矩形的长和宽之比,最小外接矩形的计算方法是在 90° 范围内以一定的角度增量旋转目标区域,每次旋转之后获取其在水平垂直坐标轴方向的外接矩形并计算出该外接矩形的面积。当旋转到外接矩形的面积最小时,此外接矩形就是最小外接矩形,该矩形的长度和宽度就是目标区域的长轴和短轴,长宽比也就是长轴与短轴之比。膜状、条带状的长宽比较大,所以该特征对区分这两类有较好的效果。

(4) 最小 Feret 直径

目标区域每旋转 15° 在水平轴上的投影为 Feret

直径,当该直径最小时就是最小 Feret 直径,当该直径最大时就是最大 Feret 直径。

(5) Feret 比

Feret 比就是最大 Feret 直径与最小 Feret 直径之比。因为膜状的最小 Feret 直径在这几类中最小,且其 Feret 比较大,所以最小 Feret 直径与 Feret 比这两项特征参数对膜状有较好的区分效果。

(6) 面积

在图像中目标区域的面积就是目标边界以及内部像素的总数,孤岛状、膜状面积较小,条带状、网状、凝块状面积较大,所以面积参数对于区分类别有一定的效果。

(7) 周长

目标区域的周长在图像中指的是围成该区域的边界的长度,也即边界像素点的个数。在剩余油形态识别中主要关注目标区域的外边界周长,而忽略内部的孔洞边界。根据目标区域周长的定义以及上述获取的目标的相关数据可以得到周长的计算公式如式(3)所示,其中 $p_k(m,n)$ 为单个边界像素点长度且值为1。

$$P = \sum_{k=1}^M p_k(m,n) \quad (3)$$

表1是笔者收集的样本集的每一项特征参数的所在区间,从中能观察到这些参数对于剩余油的区分情况,表中数据为收集到的每类剩余油形态的每种参数最小值和最大值。

由表1可以看出,孤岛状的圆形度在 $0.6032 \sim 1.0000$ 之间,而其他类圆形度参数最大为 0.5787 ,所以孤岛状可用圆形度阈值识别,实验证明这样分离孤岛状正确率很高。由表1还可以看出,其他各类的各参数区间都有部分重合,除了孤岛状以外,其他类用参数设定阈值得到的结果会有误差。KNN 算法可以很好地避免这点,由于 KNN 算法主要靠周围有限邻近的样本,而不是靠判别类域的方法来确定所属的类别,因此对于类域的交叉或重叠较多的待分类样本集来说,KNN 方法较其他方法更为适合。

3 KNN 算法识别剩余油

收集样本集剩余油上述7项特征参数,将这些已经贴好类标签的数据作为样本集特征向量空间。通过将待分类样本集特征数据和这些特征参数做归一化处理,消除不同量纲带来的影响,然后求得

表 1 参数最值

| 类别 | | 圆形度 | 凹度 | 长宽比 | 周长 | 面积 | Feret 比 | 最小 Feret |
|-----|-----|---------|---------|---------|-------|--------|---------|----------|
| 孤岛状 | 最小值 | 0.603 2 | 0.965 4 | 1.102 0 | 26 | 52 | 1.093 6 | 7.134 6 |
| | 最大值 | 1.000 0 | 1.075 2 | 2.131 2 | 115 | 470 | 2.561 4 | 25.945 |
| 凝块状 | 最小值 | 0.109 4 | 0.998 2 | 1.328 2 | 70 | 168 | 1.283 1 | 9.762 6 |
| | 最大值 | 0.578 7 | 1.266 5 | 6.491 3 | 385 | 1 312 | 4.411 8 | 55.507 |
| 膜状 | 最小值 | 0.039 7 | 1.058 5 | 3.472 6 | 65 | 63 | 1.347 1 | 5.552 1 |
| | 最大值 | 0.189 7 | 1.281 6 | 13.657 | 290 | 397 | 6.421 8 | 48.367 |
| 条带状 | 最小值 | 0.022 8 | 1.277 5 | 2.108 0 | 354 | 1 008 | 2.150 9 | 37.390 |
| | 最大值 | 0.092 5 | 1.769 7 | 4.659 0 | 1 222 | 5 563 | 4.567 0 | 114.08 |
| 网状 | 最小值 | 0.010 7 | 1.378 3 | 1.132 5 | 619 | 1 739 | 1.259 3 | 72.852 |
| | 最大值 | 0.079 3 | 2.190 9 | 2.275 6 | 4 334 | 15 933 | 2.303 4 | 341.32 |

欧氏距离,待分类剩余油近邻中哪个类别最多就判为哪类^[9-12]。

由于孤岛状比较特殊,其可通过圆形度直接分离。判为孤岛状的阈值可用下式计算得出:

$$T = \frac{\min(A_1) + \max(A_2, A_3, A_4, A_5)}{2} \quad (4)$$

其中, T 为阈值, A_1 代表孤岛状最小圆形度, A_2, A_3, A_4, A_5 代表其他四类的最大圆形度数值。因而,凡是圆形度大于 T 的为孤岛状,小于 T 的则用 KNN 算法来分类。

因为各个特征数据的量纲不同,如表 1 所示的圆形度参数最大为 1,而面积参数最大能到 15 933。如果直接使用这些未经处理的数据,则分类结果会被这些量纲大的参数(如面积等)决定,而量纲小的参数(如圆形度等)对结果的影响微乎其微。因此各特征参数不能直接用于 KNN 算法分类计算,要先将样本集和待分类样本集所有数据做归一化处理^[10]。归一化公式如式(5)所示:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (5)$$

以面积为例, X 为原面积值, X_{\max} 为样本集和待分类样本集中最大面积值, X_{\min} 为样本集和待分类样本集中最小面积值, X' 为归一化后的面积值。 X' 的范围在 0~1 之间,当各个参数经过归一化之后,就能消除量纲不同所带来的影响。

用样本集和待分类样本集归一化后的数据求解欧氏距离,计算公式如式(6)所示:

$$d_{12} = \sqrt{\sum_{k=1}^n (x_{1k} - x_{2k})^2} \quad (6)$$

其中, d_{12} 为所求欧氏距离, x_{1k} 为待分类剩余油数据, x_{2k} 为样本集数据。得到待分类剩余油与样本集的所有欧氏距离后,根据最近的 K 个近邻投票,将其判为票数最多的类。 K 值是由实验得出的,太大或太小都影响实验结果。

4 剩余油形态识别结果分析

待分类图像经过上节介绍的步骤进行分类之后,分类结果如图 3 所示,图 4 为类颜色标识。分类之后 5 个类都用不同颜色进行标注,从图 3 可以看出,数量占比最多的是凝块状,其次是孤岛状,网状、条带状、膜状数量较少。因实验用图是水驱油图,所以膜状剩余油数量较少,只有图 3(b)中有。

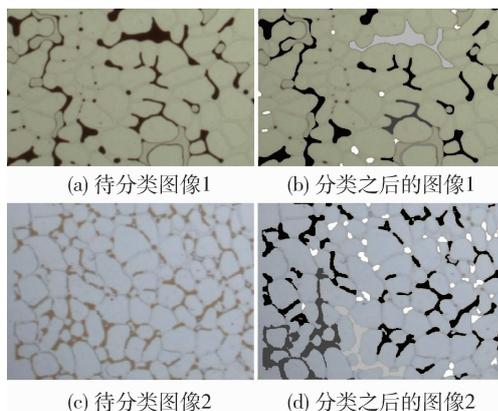


图 3 原始图和分类图



图 4 类颜色标识

实验对 5 张图像进行测试,测试分类结果如表 2 和表 3 所示,表 3 是所有测试实验图像 5 个类别的具体分类结果。

表 2 总体分类结果

| 实验图像 | 数量 | 识别正确数 | 正确率/% |
|------|-----|-------|-------|
| 1 | 135 | 122 | 90.4 |
| 2 | 112 | 103 | 91.9 |
| 3 | 127 | 114 | 89.8 |
| 4 | 124 | 112 | 90.3 |
| 5 | 141 | 129 | 91.4 |
| 总 | 639 | 580 | 90.8 |

表 3 实验图像各个类分类结果统计

| 类别 | 数量 | 分类正确数量 | 错误分类类别 | 正确率/% |
|-----|-----|--------|--------|-------|
| 孤岛状 | 274 | 253 | 凝块状 | 92.3 |
| 凝块状 | 236 | 213 | 不定 | 90.3 |
| 条带状 | 53 | 48 | 网状 | 90.5 |
| 膜状 | 35 | 31 | 凝块状 | 88.6 |
| 网状 | 41 | 36 | 条带状 | 87.8 |

根据表 2 的每张图像统计结果来看,KNN 算法用于剩余油分类有良好的分类结果。从表 3 能看出孤岛状用圆形成度阈值分类效果很好,KNN 算法对剩余 4 类均有较好的识别率。膜状的错误识别是因为有一些狭长形状的凝块状,其与膜状相似,所以就有可能将膜状误识别为凝块状。网状和条带状的相互误识别在于条带状的最小 Feret 直径较大时和网状较为相似,同理网状的最小 Feret 直径较小时与条带状相似,其根本原因在于网状和条带状的界限不够明显。

5 结论

因为剩余油的分类类别没有严格的界限,各类的数据都有交叉,所以导致对类的识别有一些模糊性和不确定性。而使用 KNN 算法的好处在于,能够在一定程度上消除特征参数相互交叉带来的影响,用近邻投票选出类别,实验结果证明效果良好。通过观察大量的剩余油特征数据,挑选出对识别剩余油类别有一定作用的参数,剔除一些无用参数,使得这些参数在识别剩余油类别时更加高效。再通过数据对比用圆形成度将孤岛状分离出去,减少了一部分算法分类的时间,也能够保证孤岛状分类的准确性。用 KNN 算法直接对剩余油进行形态识别,既

能减少大量人工的工作量,又能保证识别的正确率,具有比较高的应用价值。

参考文献

- [1] 徐清华. 大庆油田三元复合驱后微观剩余油分布特征[J]. 大庆石油地质与开发,2019,38(4):110-116.
- [2] 元征仁,刘新. 剩余油技术发展历程[J]. 石油知识,2018(2):10-11.
- [3] LI J J, LIU Y, GAO Y J, et al. Effects of microscopic pore structure heterogeneity on the distribution and morphology of remaining oil[J]. Petroleum Exploration and Development,2018,45(6):1112-1122.
- [4] Li Y B. Analysis of remaining oil distribution and influencing factors in B development zone[C]. IOP Conference Series:Earth and Environmental Science,2019.
- [5] 佟颖,贾元元. 基于刻蚀模型的高渗条带控制下剩余油微观赋存特征[J]. 科学技术与工程,2018,18(28):80-86.
- [6] 崔嘉祯,刘人铜. 剩余油分布影响因素及数值模拟方法的应用[J]. 云南化工,2019,46(2):171-172.
- [7] 李洪生. 双河油田聚合物驱后微观剩余油分布特征[J]. 西安石油大学学报(自然科学版),2018,33(3):70-74.
- [8] 陈浩,滕奇志,何小海,等. 基于几何形状特征的剩余油形态识别[J]. 微型机与应用,2017,36(1):18-21.
- [9] 胡君萍,傅科学. 基于改进 KNN 算法的手写数字识别研究[J]. 武汉理工大学学报(信息与管理工程版),2019,41(1):22-26.
- [10] DERRAC J, CHICLANA F, GARCÍA S, et al. Evolutionary fuzzy k-nearest neighbors algorithm using interval-valued fuzzy sets[J]. Information Sciences, 2016, 329: 144-163.
- [11] JIANG J H, CHEN Y J, MENG X Q, et al. A novel density peaks clustering algorithm based on k nearest neighbors for improving assignment process[J]. Physica A: Statistical Mechanics and its Applications,2019,523:702-713.
- [12] 桑国通,廖晓曦,何欣龙,等. K 近邻算法结合红外光谱对轮胎橡胶颗粒的鉴别研究[J]. 化学通报,2019,82(1):87-91.

(收稿日期:2019-10-15)

作者简介:

程小龙(1994-),男,硕士研究生,主要研究方向:图像处理与数字通信。

王正勇(1969-),女,博士,副教授,硕士生导师,主要研究方向:图像处理与模式识别。

滕奇志(1962-),女,博士,教授,博士生导师,主要研究方向:图像处理与模式识别。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部
中国电子信息产业集团有限公司第六研究所