

doi: 10.11720/wtyht.2020.0069

蔡泽园,鲁宝亮,熊盛青,等.基于自适应核密度的贝叶斯概率模型岩性识别方法研究[J].物探与化探,2020,44(4):919-927.http://doi.org/10.11720/wtyht.2020.0069

Cai Z Y, Lu B L, Xiong S Q, et al. Lithology identification based on Bayesian probability using adaptive kernel density[J]. Geophysical and Geochemical Exploration, 2020, 44(4): 919-927. http://doi.org/10.11720/wtyht.2020.0069

基于自适应核密度的贝叶斯概率模型岩性识别方法研究

蔡泽园^{1,2,3}, 鲁宝亮^{1,2,3}, 熊盛青⁴, 王万银^{1,2,3}

(1. 长安大学 重磁方法技术研究所, 陕西 西安 710054; 2. 长安大学 地质工程与测绘学院, 陕西 西安 710054; 3. 长安大学 西部矿产资源地质工程教育部重点实验室, 陕西 西安 710054; 4. 中国自然资源航空物探遥感中心, 北京 100083)

摘要: 准确地刻画岩石类型及其结构关系可以为能源矿产勘探、深部结构与构造等研究提供重要信息。目前利用地球物理技术可以通过不同岩石对应的物性参数(如密度、磁化率、电阻率、速度等)之间的差异进行岩性识别,但是不同岩石物性往往存在一定程度的重合,利用单一物性进行岩性识别的结果不够准确,因此利用多源数据进行岩性识别具有重要的意义。贝叶斯方法属于统计分类方法,依靠概率进行分类,概率密度的计算依靠样本属性之间的相互联系。基于此,我们将基于自适应核密度估计的贝叶斯概率模型引入到岩性识别中。该方法对于多类不同物性参数具有良好的适应能力,预测的岩性分类结果带有概率参数,可以存在模糊区间,提供多种岩性分类结果。该方法具有较强可扩展性,可以同时处理参数和非参数信息,使得已知地质信息以及物性参数得到最大化的利用。实验证明该方法的岩性识别结果较好,相比于传统高斯算法和固定带宽核密度估计,自适应带宽的核密度估计获得的分类结果更稳定、更准确。

关键词: 贝叶斯; 岩性识别; 物性参数; 自适应核密度

中图分类号: P631

文献标识码: A

文章编号: 1000-8918(2020)04-0919-09



0 引言

岩性识别是地质研究过程中非常重要的基础工作,尤其是在近地表以及深部无法直接采样区的地质研究中,准确地刻画深部岩石类型及其结构关系,可以为能源矿产勘探、深部结构与构造等研究提供重要的地质信息。因此采用什么数据、什么方法进行岩性识别是一项极具价值的研究工作。传统(地表)地质填图已远不能满足深部探测的需要。随着多源地学数据(地质、地球物理、地球化学、遥感以及钻井数据等)的获取、地球物理三维反演技术和三维地质建模的发展,深部岩性识别已成为现

实。但目前对多源数据的利用并不充分,尚未有效地利用多种物性参数进行岩性识别。因此,如何利用多源不同类型、属性地学数据所反映的岩性信息,从高维数据空间准确地进行岩性识别是亟待解决的难题,对深部探测和地学数据融合也具有重要的科学意义。

目前岩性识别主要有直接手段(岩芯、手标本以及薄片鉴定等)和间接手段(重磁、地震、电磁、测井、遥感、地球化学等)。从深部探测来讲,钻井岩芯可能是唯一的岩石标本,但只是点数据,而遥感(面数据)只能探测地表情况,且受地表覆盖影响大;因此地球物理方法在地下深部岩性识别研究中将发挥重要作用。而物性(密度、磁化率、电导率、

收稿日期: 2020-02-17; 修回日期: 2020-05-28

基金项目: 国家重点研发计划项目“典型覆盖区航空地球物理技术示范与处理解释软件平台开发”(2017YFC0602200)之课题“航空地球物理数据综合处理解释方法研究及软件开发”(2017YFC0602202);国家自然科学基金项目(41904106)

作者简介: 蔡泽园(1996-),女,长安大学硕士研究生,主要从事重磁数据处理与解释研究工作。Email: 847330992@qq.com

通讯作者: 鲁宝亮(1984-),男,博士,副教授,主要从事重磁位场反演与深部构造研究工作。Email: lulb@chd.edu.cn

纵横波速度等)是岩性和地球物理场之间的纽带,因此,通过地球物理反演物性,再联合其他地质数据进行岩性的识别具有可行性。地震反演是进行岩性识别的有效方法,可以利用地震不同的弹性参数,如利用纵横波速度、密度^[1-2]、波阻抗、振幅、频率、相位^[3-5]对目标岩体(流体)进行识别,但是当岩相间的地震响应差别不明显时,依靠波形的分类结果无法准确刻画岩性且地震方法成本高。在测井识别岩性方面,可以利用交会图版^[6-9]或者基于统计、聚类、支持向量机以及神经网络等技术^[10-14]进行岩性识别,测井方法在精度、算法以及技术上都有明显的优势,但是只能识别钻孔附近小范围内的岩性,难以进行大面积或是没有钻孔地区的岩性识别。利用重磁技术识别岩性主要依靠密度与磁化率比值,例如采用交会图版结合逻辑运算识别岩性^[15-16],重磁数据覆盖面积广,采样密度高,容易获得大规模的岩性识别,但是在无约束情况下,垂向分辨较差,多解性强。

单一的地球物理方法往往很难获得理想识别结果,因此联合多个地球物理方法的岩性识别成为主流方式。在多源数据识别岩性方面,对于存在岩性与物性对应关系的模糊区域,以地震反射特征为约束,利用重磁电资料识别了具有密度、磁性和电阻率等特征差异的火成岩岩性^[17]。近年来随着机器学习(machine learning)的兴起,该算法已被广泛应用于岩性识别。如利用多种地球物理数据基于模糊聚类分析^[18]将岩石进行分类识别。利用地球化学数据和地球物理数据对岩性进行多元回归分析^[19-20]或者多准则决策方法识别。利用无监督模糊分区聚类对航空伽马射线数据与陆地卫星波段数据联合进行岩性识别^[21]。应用受限玻尔兹曼机(restricted boltzmann machine)和随机森林模型到区域尺度多参数地球科学数据集,从而预测斑岩铜金矿床的远景区域。还有采用随机森林法(random forests)和自我组织映射技术(SOM, self-organising maps)来识别连续的火山单元子类,由于算法只针对部分地球科学或地质参数进行岩性识别,因此无法针对不同类型的参数进行统一处理。虽然各类算法都有改进,但是岩性识别结果唯一,对模糊区间的多种可能性无法准确表述。

理论上讲,通过不同地质、地球物理技术可以获得地下物性结构,如密度结构、速度结构、电阻率结构、磁性结构等,那么如何将这些物性结构转换为岩性是值得研究的问题,实质上这是一个模式识别问题。通常来讲,岩性与物性的对应关系并不总是明

确的,在交会图上存在较大的重叠区域,从而使得基于规则的分类方法难以解决该问题。在前人的研究基础之上,考虑到贝叶斯方法是非规则分类,该方法依据类的概率、概率密度,并按照某种规则使得分类结果从统计上达到最佳。基于此,笔者提出了基于自适应核密度估计的贝叶斯概率岩性识别方法,完成了从物性到岩性的转换。该方法具有较强的泛化能力,预测的岩性分类结果带有概率参数,可以存在模糊区间,提供多种岩性分类结果。该方法具有较强可扩展性,可以有任意数量类型的输入参数(允许存在缺省参数)以及任意数量的岩性分类输出。通过实验对比了传统的高斯密度、固定带宽核密度以及自适应带宽核密度对岩石物性数据判别的效果,说明了该方法具有良好岩性识别效果。

1 基本原理

1.1 贝叶斯分类

贝叶斯算法是基于统计学的基本算法之一,假设各个条件之间相互独立,可以得到朴素贝叶斯算法。如果我们将岩石的类型表示为 c 事件,将岩石属性表示为 x 事件,岩石类型 c 和对应属性 x 是发生在同一空间的两个事件,假设某研究区的完整岩石类型 c 是由两种岩石类型 c_1, c_2, \dots, c_n 构成, c_1, c_2, \dots, c_n 中一个岩石种类出现必然伴随着某一属性 x 的发生,即若 x 发生,则 c 必然有一个会发生,根据概率可以得到样本集的已知各类别 c_i 的先验概率以及各类条件概率 $P(x/c_i)$ 。对于未知样本,贝叶斯公式可以计算出待测样本分属各类的概率,称为后验概率。

使用贝叶斯定理来预测后验概率最大的类,主要是估计每一类的概率密度函数,通过多元正态分布来建模。朴素贝叶斯分类器基于条件独立性假设,是概率分类器中最简单的分类器,在很多情况下具有相当高的分类准确率,因此以高效率和良好的泛化能力而著称。对于某个测区,已知该地区的物性分布特征(如密度、磁化率、电阻率等)以及部分的岩石样本,假设属性之间相互独立,可以使用贝叶斯分类器来对整个测区的岩石类型进行预测。概率分类器的优点在于在得到分类结果的同时,会对每一种类别进行概率计算,对于岩性的识别而言,可以通过概率或相对概率来人为判定分类结果的可信度而不仅仅依靠算法本身的置信度来决定,小概率区间会提供一个模糊带,模糊带的类别区分结果可能不唯一。

贝叶斯分类器通过对每个未知样点 x 和每个类 c_i 来估计其后验概率 $P(c_i/x)$, 即计算未知样本 x 属于 c_i 类岩石的概率, 从而选择最大概率的类作为未知样本 x 最终的预测类型。利用贝叶斯定理, 后验概率 $P(c_i/x)$ 可以表示为:

$$P(c_i/x) = \frac{P(x/c_i)P(c_i)}{P(x)}, \tag{1}$$

其中: $P(x/c_i)$ 定义为假设真实类是 c_i 时观察到 x 的概率, 称为似然; $P(c_i)$ 是类 c_i 发生的先验概率; $P(x)$ 指从全部样本中观察到 x 的概率, 可以表示为 $P(x) = \sum_{i=1,k} P(x/c_i)P(c_i)$ 。

对于给定的一点 x , $P(x)$ 是确定的。用 $f_i(x)$ 表示未知样本 x 属于 c_i 类的概率密度, 似然用概率密度可以表示为 $P(x/c_i) = 2af_i(x)$, 其中 a 表示邻近 x 的一个极小区间, 因此可以得到后验概率的计算公式:

$$P(c_i/x) = \frac{P(x/c_i)P(c_i)}{P(x)} \propto P(x/c_i)P(c_i) \propto f_i(x)P(c_i), \tag{2}$$

由式(2)可知, 概率密度函数是影响分类结果的一个重要因素, 由于大部分岩石物性参数是基本遵循正态分布的规律, 这里考虑了传统的高斯公式作为概率密度函数, 然而样本本身并不完全遵循正态分布, 为了极大地避免由于选定高斯函数的影响, 同时考虑使用核密度估计的方法, 核密度估计的优点在于核函数的选取对于最终分类结果影响不大, 更适用于类正态数据样本, 这会在之后的模型测试中加以验证。

1.2 自适应核密度估计

核密度估计^[22-23]是统计学中用来估计随机变量的概率密度函数的方法, 属于非参数检验方法的一种。核密度估计的基本表达式为:

$$f_i(x) = \frac{1}{nh_i} \sum_{i=1,n} K\left(\frac{x - X_i}{h}\right), \tag{3}$$

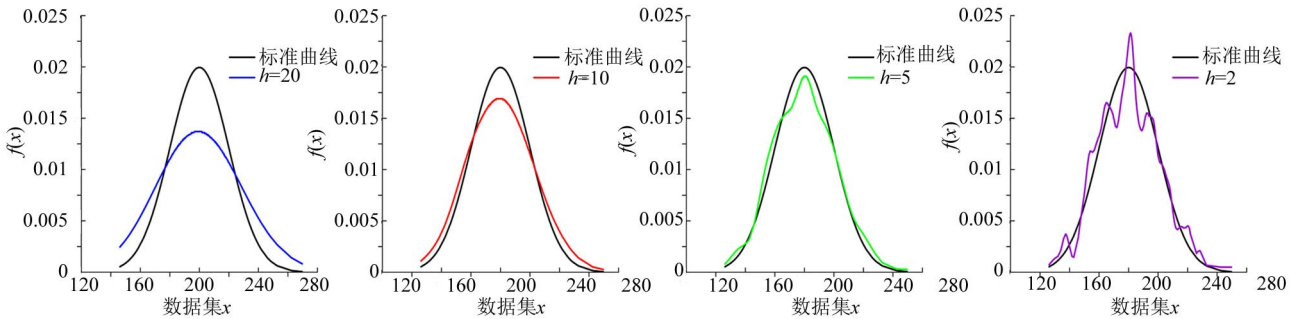


图1 不同带宽下概率密度函数曲线

Fig.1 Probability density function curve under different bandwidth

其中: n 表示所有已知样本的总个数, h 表示带宽, 也可以称为窗宽或滑动参数, x 表示随机样本, X_i 表示第 i 个已知样本, $K(\cdot)$ 表示核函数。核函数的作用相当于权函数, 可以根据距离分配各个样本点对总体密度估计贡献的不同程度, 常见的有高斯核函数, 三角核函数, 二次核函数等多种表达形式, 由于大部分的物性分布都遵循正态分布的原则, 因此这里采用高斯核函数, 即 $K(t) = e^{-0.5t^2} / \sqrt{2\pi}$ 。值得一提的是, 在大样本量的条件下, 核密度估计中核函数 K 的选取对于概率密度的影响是很小的, 可以忽略的^[24]。

核密度带宽的选择是得到最佳估计结果的关键。固定带宽是比较常见的方法^[25-26]。不同的带宽选取, 会对概率密度函数产生不一样的影响, 由于概率密度函数存在一个必然性, 即概率密度之和为 1, 因此, 不同的带宽在影响曲线光滑程度的同时, 也会对峰值产生影响, 曲线越平滑, 峰值就会越低, 曲线越抖, 峰值就会越高。带宽选择的基本思想是基于最小平方差, 根据积分均方误差最小, 求出最优带宽。

图 1 展示了对于 1200 组完全正态分布的样本, 选取不同带宽得到的概率密度函数曲线图, 可以得到, 对于该组数据而言, 下列所选带宽的结果拟合程度最高的是 $h = 5$ 时, 当所选带宽 h 过大, 会造成核密度估计曲线过于平滑, 从而失去应有的特征细节。带宽 h 过小, 会导致核密度估计曲线光滑性差, 过于粗糙, 会产生过拟合的问题。对于高斯核函数, 最优带宽为:

$$h_{opt} = 1.059\sigma n^{-0.2}. \tag{4}$$

岩石的物性参数往往存在较大的差异, 而每种参数的误差范围也不同, 因此实际操作中固定带宽的方法可能并不适用于岩性的识别, 加上在实际采样中, 受限于各种地理和人为因素的影响, 无法保证样本的稀疏程度一致, 对于不同质量的样本, 无法

用同一带宽来进行密度估计,需要对不同稀疏度的样本分别讨论带宽,因此相比于固定带宽法,滑动变带宽更能满足岩性预测的要求。令带宽值随着数据的密度变化自动进行适当的调整,实现滑动变带宽的方法,主要是基于积分均方误差,通过计算每个点的最优带宽值,得到变带宽函数 $h(x)$ 。针对滑动窗口的实现过程,于传强等^[27]给出了详细的推导过程,关于滑动变带宽的算法流程如下:

第一,根据固定带宽的经验公式,选择样本组的最优固定带宽 h_{opt} 以及对应的核密度估计函数 $f_{opt}(x)$,

$$f_{opt}(x) = \frac{1}{nh_{opt}} \sum_{i=1,n} K\left(\frac{x - X_i}{h_{opt}}\right)。$$

(5)

第二, 根据给定的估计点,求取优化后的带宽

$$h(x) = h_{opt} c^{0.2} \frac{f_{opt}(x)^{0.2}}{\left[\sum_{i=1,n} \frac{1}{nh_{opt}} t^2 K(t) - f_{opt}(x)\right]^{0.4}},$$

(6)

其中 $c=0.375\pi^{-0.5}h_{opt}^{-5}$ 。

第三,优化后的核密度估计函数记为

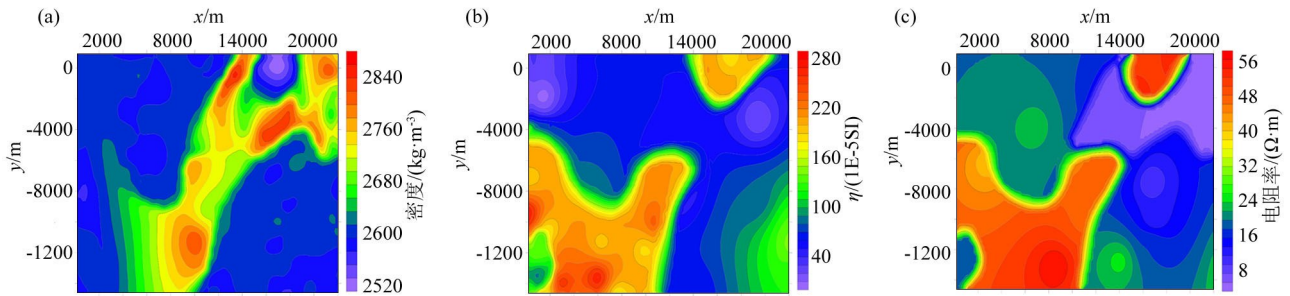
$$f(x) = \frac{1}{nh(x)} \sum_{i=1,n} K\left(\frac{x - X_i}{h(x)}\right)。$$

(7)

上述计算的带宽在数据质量相对较好的情况下往往会出现过拟合现象,反而结果不尽如人意,因此在实际的分类过程中,当计算得到固定带宽以及优化后的带宽后,需要对带宽的差值进行判定,对于二者相差不大的情况(差值需要根据样本的数量和质量人为给定),采用优化前的带宽。将优化后的核密度估计函数作为贝叶斯模型的概率密度函数,针对某一未知样本,可以得到该样本归属于每一类的概率密度,通过比较得到最大概率类别作为预测的类。

2 模型测试

本次实验选取密度、磁化率和电阻率作为分类依据,测区内岩石类型为板岩、片麻岩和花岗岩,模型具体参数见表 1,从表中可以看出,整个模型的密度变化较小,但是相比于其他两种岩石的密度,花岗岩的密度范围变化更小,主要集中在 $2\,600\text{ kg/m}^3$ 左右,可以作为划分花岗岩的物性依据。图 2、3 分别是模型区内的岩石物性分布情况以及该模型下的岩石分布示意图,为了对比分类器模型的准确率、稳定性以及计算复杂度,从所有的样本中选取不同位置、



a—模型密度;b—模型磁化率;c—模型电阻率
a— density of the model;b—magnetic susceptibility of the model;c—resistivity of the model

图 2 模型物性

Fig.2 Physical properties of the model

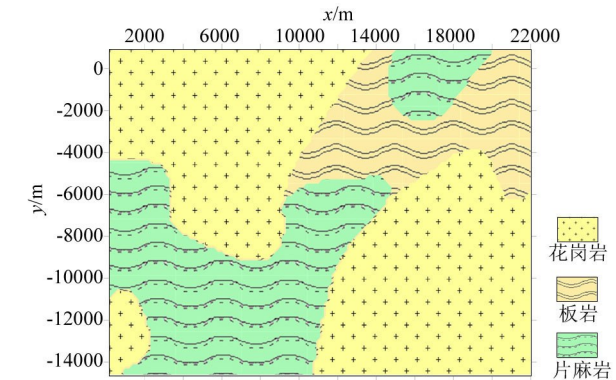


图 3 岩石分布

Fig.3 rock distribution of the model

不同数据量的样本作为训练集和测试集进行分类,判定分类结果。

该模型共有 8 690 个样本点,分别选取 250、435、870 个点作为训练样本,其他点作为测试样本,图 4~9 分别为不同训练样本的物性参数交互图以及分类结果,表 2 是关于不同训练样本错误率统计表。

对照红色散点和灰色的岩性边界可以看出,识别错误的样本集中分布在边界的低概率区,而岩性边界可以看作整个区域的模糊区间,因此概率密度图在低概率区刻画了整个区域的模糊区,对比图 5、

表 1 模型参数统计

Table 1 Statistical table of model parameters

岩石名称	密度/($\text{kg} \cdot \text{m}^{-3}$)	磁化率/(10^{-5}SI)	电阻率/($\Omega \cdot \text{m}$)
板 岩	2630~2850	0~160	3~8
片麻岩	2570~2830	180~280	40~60
花岗岩	2580~2640	0~160	10~30

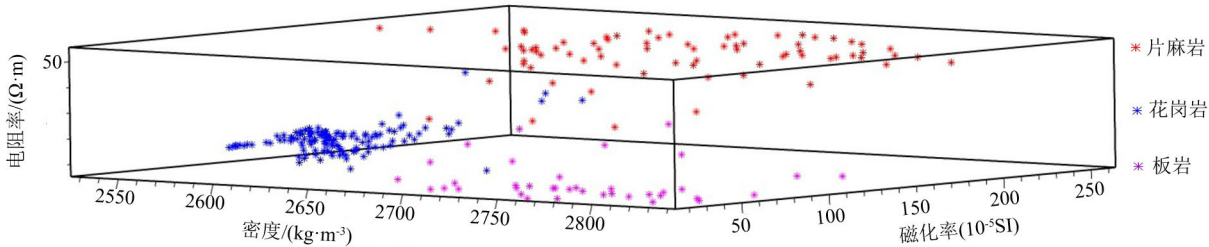
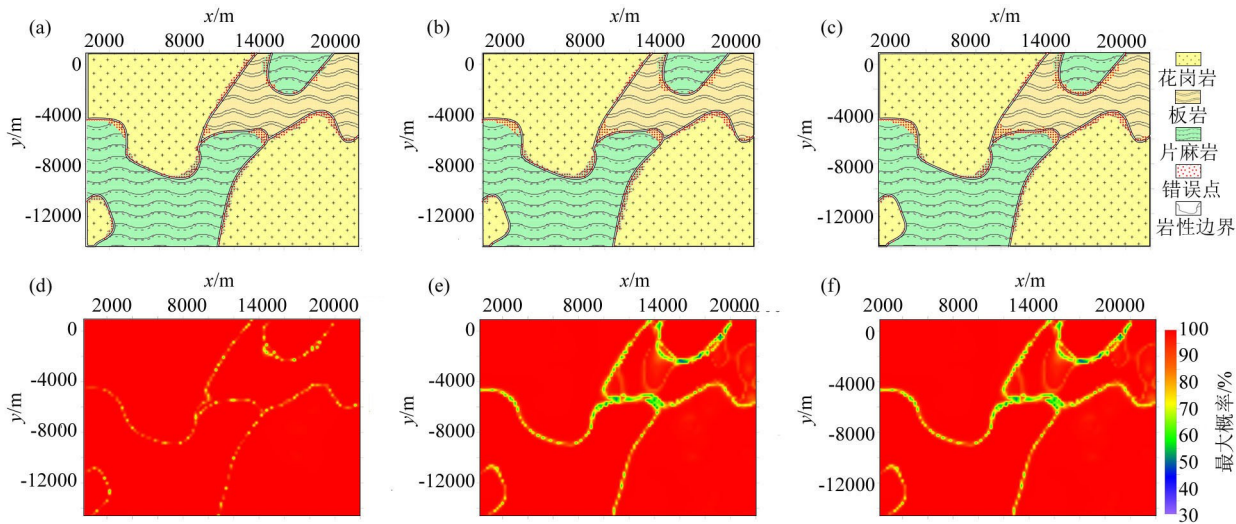


图 4 250 点训练样本物性参数交互图

Fig.4 Interaction diagram of physical parameters of 250 training samples



a~f 分别表示对于 250 个训练样本点的传统高斯分类、固定带宽的核密度估计、自适应带宽的核密度估计的贝叶斯分类结果以及其对应的概率分布

Figures a~f respectively represent the Bayesian classification results, corresponding probability distribution map of the traditional Gaussian classification, fixed bandwidth kernel density estimation, adaptive bandwidth kernel density estimation for 250 training sample points

图 5 250 点训练样本分类结果

Fig.5 Classification results of 250 training samples

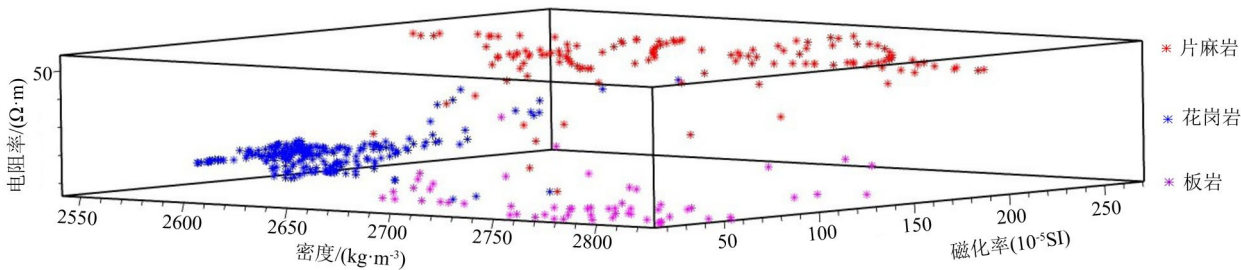
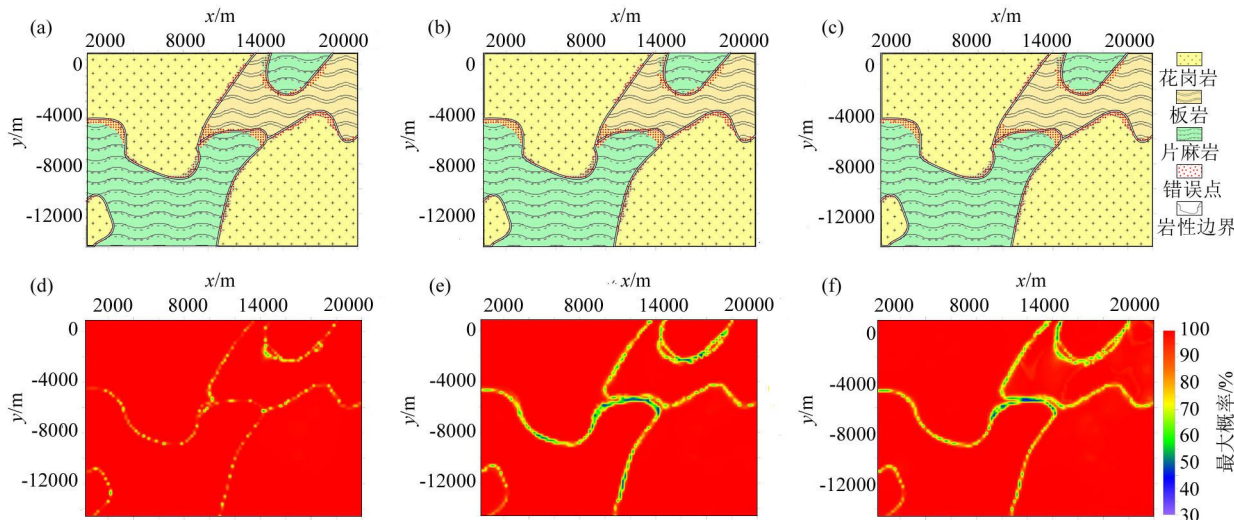


图 6 435 点训练样本物性参数交互图

Fig.6 Interaction diagram of physical parameters of 435 training samples

表 2 模型错误率统计
Table 2 Statistical table of model error rate

训练样本/个	错误率/%		
	传统高斯算法	固定带宽核密度估计	自适应带宽核密度估计
250	4.52	4.48	4.17
435	4.58	4.51	4.09
870	4.46	4.37	4.20



a~f 分别表示对于 435 个训练样本点的传统高斯分类、固定带宽的核密度估计、自适应带宽的核密度估计的贝叶斯分类结果以及其对应的概率分布

Figures a~f respectively represent the Bayesian classification results, corresponding probability distribution map of the traditional Gaussian classification, fixed bandwidth kernel density estimation, adaptive bandwidth kernel density estimation for 435 training sample points

图 7 435 点训练样本分类结果
Fig.7 Classification results of 435 training samples

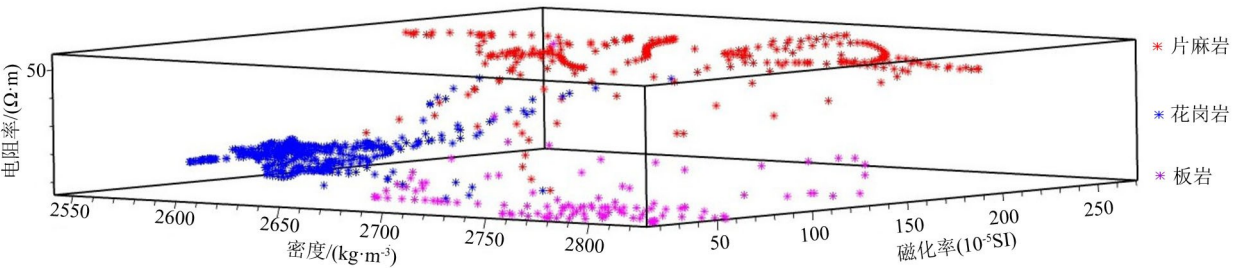


图 8 869 点训练样本物性参数交互图

Fig.8 Interaction diagram of physical parameters of 869 training samples

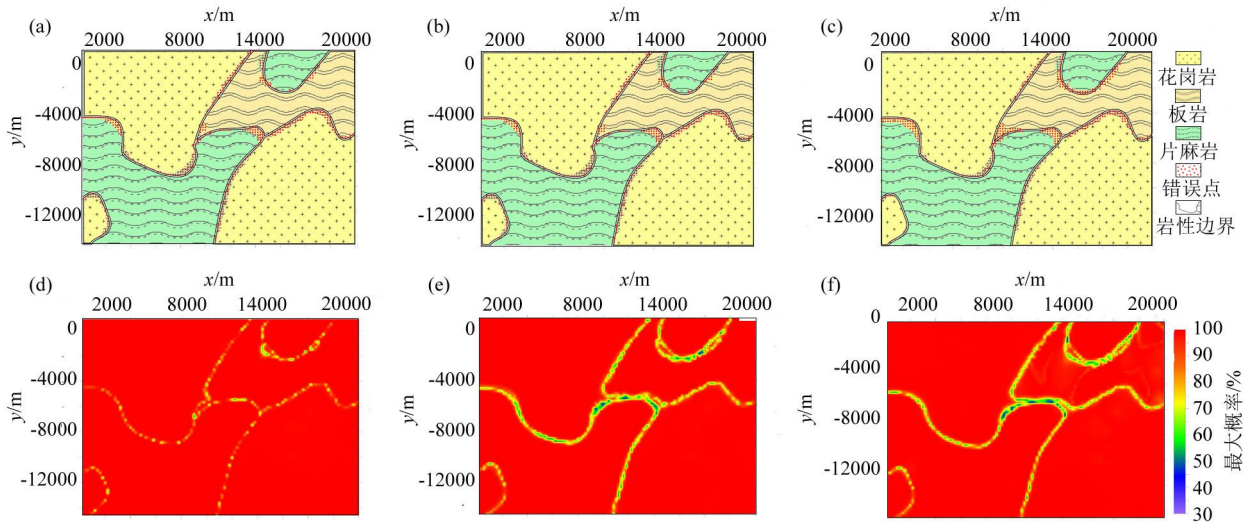
图 7、图 9 的概率图, 针对同一训练样本, 核密度估计算法对于模糊区的刻画整体优于传统的高斯算法估计模型; 对于同一算法下的不同训练样本, 传统高斯算法的低密度带并没有表现出由于训练样本增加其结果得到改善, 而随着训练样本的增加, 自适应带宽核密度估计算法的概率图上对于分类结果正确区域的低概率带有了明显改善。

从表 2 的错误率上而言, 在同样的训练样本下, 基于自适应核密度估计的贝叶斯概率模型明显优于其他两类模型。由此可知, 贝叶斯概率密度模型对

于深部岩性识别的方法是有效可行的, 而基于自适应核密度估计的贝叶斯概率模型相比于其他概率密度模型而言效果相对较好。

通过测试不同训练样本对于分类结果的影响, 最终可以得到针对该模型, 训练样本每类都少于 40 个时便无法进行合理的岩性识别, 当然并非训练样本的数量越多越好, 样本质量对于识别结果也有着决定性作用。

通过对比 3 种概率密度方法得到的分类结果可知, 基于自适应带宽的贝叶斯概率模型下的分类错



a~f 分别表示对于 869 个训练样本点的传统高斯分类、固定带宽的核密度估计、自适应带宽的核密度估计的贝叶斯分类结果以及其对应的概率分布

Figures a~f respectively represent the Bayesian classification results, corresponding probability distribution map of the traditional Gaussian classification, fixed bandwidth kernel density estimation, adaptive bandwidth kernel density estimation for 869 training sample points

图 9 869 点训练样本分类结果

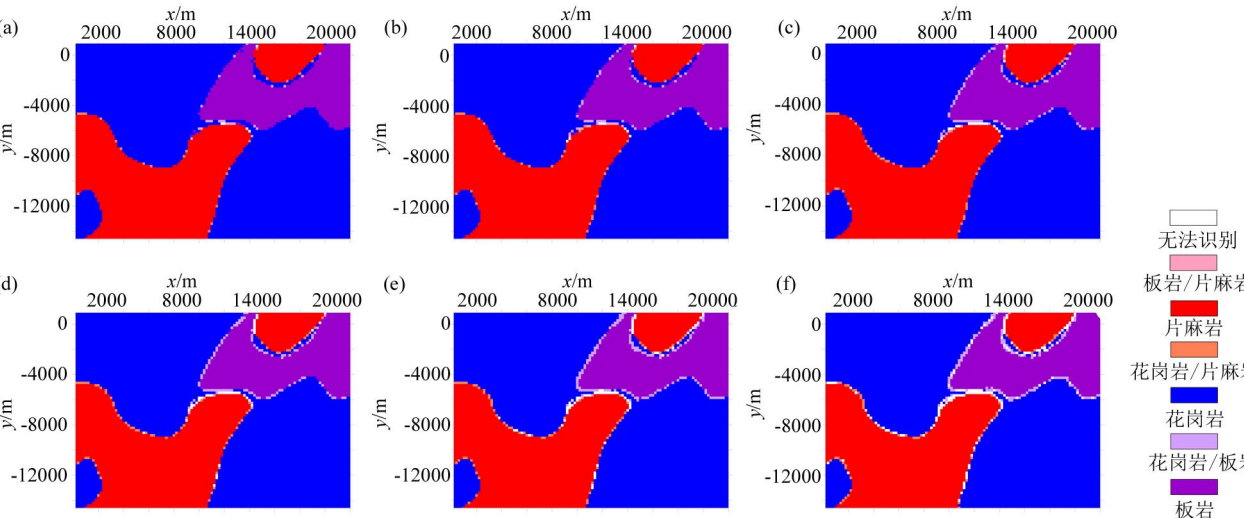
Fig.9 Classification results of 435 training samples

误率是相对较低的且在模糊区有一个明显的低概率带,因此,模型二将针对模型一中 435 点训练样本在自适应带宽下的测试样本进行概率差下的统计分析。

若两种类型岩石的预测概率差在拟定的差值之内,就认为该未知样本属于这两类岩性的概率相同,对于 3 种岩性也同样适用,对于该模型而言,若 3 种岩性的概率差都在拟定的差值内,则认为无法对该样本进行岩性识别,选择更大的概率差,意味着模型

的容错率降低,同时,识别类型唯一的样本可信度也随之变大,针对不同的概率差,预测结果不同。

图 10 为概率差在 10%、20%、30%、40%、50%、60% 概率差下的岩性预测结果,白色区域表示 3 种岩性被认为是等概率的情况,即无法识别的区域,同时,浅色区域代表可能为两种岩性。在无法进行准确的岩性识别的区域进行人工识别或者二次识别,避免了机器学习在岩性识别过程中由于算法本身的局限性,导致预测结果唯一且无法进行人工推断可



a~f 依次为概率差在 10%、20%、30%、40%、50%、60% 时的分类结果

a~f are the classification results when the probability difference is 10%, 20%, 30%, 40%, 50% and 60% respectively

图 10 预测分类结果

Fig.10 Forecast classification results

信度的问题,在机器学习和人工识别中找寻一个平衡,发挥二者优势,在提高岩性识别效率和准确度的同时使得预测结果明朗可操作。

3 结论与建议

笔者将基于自适应核密度估计的贝叶斯概率模型应用到岩性识别中,该方法的优势在于,对于有一定重合区域的物性参数,可以有效地进行未知区域的岩性识别,提高计算的精度和效率,将多种物性参数进行合理的处理解释,提高数据利用率,通过各个数据之间的相互作用,最终获得岩性识别结果。

从本文的模型可以看出,基于自适应核密度估计的贝叶斯概率模型能够较好地刻画未知地区岩性的类别和轮廓,对于模糊区也可以较好地进行判定。事实上,该方法并不局限于以上几种参数,对于其他的物性参数也可以进行同样的处理,识别结果的精度也依靠于更多类型的物性参数和更好质量的训练样本。

参考文献 (References):

- [1] Lortzer G J M, 杨谦, 译. 岩性反演的完整方法第一部分: 理论 [J]. 国外油气勘探, 1993, 5(4): 414–428.
Lortzer G J M, Yang Q, Trans. Complete lithology inversion method of the first part theory [J]. Foreign Oil and Gas Exploration, 1993, 5(4): 414–428.
- [2] 田玉昆, 周辉, 袁三一. 基于马尔科夫随机场的岩性识别方法 [J]. 地球物理学报, 2013, 56(4): 1360–1368.
Tian Y K, Zhou H, Yuan S Y. Lithologic discrimination method based on Markov random-field [J]. Chinese J. Geophysics, 2013, 56(4): 1360–1368.
- [3] 靳军, 刘楼军, 邵雨, 等. 综合地球物理方法识别准噶尔盆地的岩性圈闭 [J]. 石油地球物理勘探, 2002, 37(3): 287–290, 299.
Jin J, Liu L J, Shao Y, et al. Discussion on identifying method for identification of lithologic traps in Junggar Basin by comprehensive geophysical method [J]. Oil Geophysical Prospecting, 2002, 37(3): 287–290, 299.
- [4] 洪忠, 张猛刚, 苏明军. 应用地震波形分类技术识别岩相的适用性和局限性 [J]. 物探与化探, 2013, 37(5): 904–910.
Hong Z, Zhang M G, Su M J. The applicability and limitations of the seismic wave-form classification technology to the identification of lithological facies [J]. Geophysical and Geochemical Exploration, 2013, 37(5): 904–910.
- [5] Tjelmeland, Håkon, Luo X, et al. A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image [J]. Geophysical Prospecting, 2019, 67(3): 609–623
- [6] 宫清顺, 黄革萍, 孟祥超, 等. 三塘湖盆地火山岩岩性识别方法 [J]. 中国石油勘探, 2012, 17(3): 37–41, 6.

- Gong Q S, Huang G P, Meng X C, et al. Methods for lithology discrimination of volcanics in Santanghu Basin [J]. China Petroleum Exploration, 2012, 17(3): 37–41, 6.
- [7] 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 D L, Li T, Huang B H, et al. Research on the identification of the lithology and fluid type of foreign Moilfield by using the cross-plot method [J]. Progress in Geophysics, 2012, 27(3): 1123–1132.
- [8] 李伟才, 姚光庆, 黄银涛, 等. 文昌 13-1 油田低阻油层测井岩性识别方法研究 [J]. 石油天然气学报, 2012, 34(12): 81–85, 7.
Li W C, Yao G Q, Huang Y T, et al. Study on identification method of logging lithology for low resistivity reservoir in Wenchang 13-1 Oilfield [J]. Journal of Oil and Gas Technology, 2012, 34(12): 81–85, 7.
- [9] 范宜仁, 黄隆基, 代诗华. 交会图技术在火山岩岩性与裂缝识别中的应用 [J]. 测井技术, 1999(1): 53–56, 64.
Fan Y R, Huang L J, Dai S H. Application of crossplot technique to the determination of lithology composition and fracture identification of igneous rock [J]. Well Logging Technology, 1999(1): 53–56, 64.
- [10] 田艳, 孙建孟, 王鑫, 等. 利用逐步法和 Fisher 判别法识别储层岩性 [J]. 勘探地球物理进展, 2010, 33(2): 126–129, 134.
Tian Y, Sun J M, Wang X, et al. Identifying reservoir lithology by step-by-step method and Fisher discriminant [J]. Progress in Geophysics, 2010, 33(2): 126–129, 134.
- [11] 王辉, 黎明碧, 唐勇, 等. 基于小波神经网络的 ODP1148A 井岩性预测 [J]. 地球物理学进展, 2014, 29(1): 392–399.
Wang H, Li M B, Tang Y, et al. Tang Y The lithology prediction of ODP hole 1148A based on the wavelet neural network [J]. Progress in Geophysics, 2014, 29(1): 392–399.
- [12] 吴施楷, 曹俊兴. 基于连续限制玻尔兹曼机的支持向量机岩性识别方法 [J]. 地球物理学进展, 2016, 31(2): 821–828.
Wu S K, Cao J X. Lithology identification method based on continuous restricted Boltzmann machine and support vector machine [J]. Progress in Geophysics, 2016, 31(2): 821–828.
- [13] 安鹏, 曹丹平. 基于深度学习的测井岩性识别方法研究与应用 [J]. 地球物理学进展, 2018, 33(3): 1029–1034.
An P, Cao D P. Research and application of logging lithology identification based on deep learning [J]. Progress in Geophysics, 2018, 33(3): 1029–1034.
- [14] Corina A N, Hovda S. Automatic lithology prediction from well logging using kernel density estimation [J]. Journal of Petroleum Science and Engineering, 2018, 170: 664–674.
- [15] 付光明, 严加永, 张昆, 等. 岩性识别技术现状与进展 [J]. 地球物理学进展, 2017, 32(1): 26–40.
Fu G M, Yan J Y, Zhang K, et al. Current status and progress of lithology identification technology [J]. Progress in Geophysics, 2017, 32(1): 26–40.
- [16] 严加永, 吕庆田, 陈向斌, 等. 基于重磁反演的三维岩性填图试验——以安徽庐枞矿集区为例 [J]. 岩石学报, 2014, 30

(4);1041–1053.

Yan J Y, Lyu Q T, Chen X B, et al. 3D lithologic mapping test based on 3D inversion of gravity and magnetic data: A case study in Lu-Zong ore concentration district, Anhui Province [J]. *Acta Petrologica Sinica*, 2014,30(4): 1041–1053.

[17] 刘云祥, 何展翔, 张碧涛, 等. 识别火成岩岩性的综合物探技术[J]. *勘探地球物理进展*, 2006,29(2): 115–118,5.

Liu Y X, He Z X, Zhang B T, et al. Integrated geophysical techniques for identification of igneous rocks [J]. *Progress in Exploration Geophysics*, 2006,29(2): 115–118,5.

[18] Paasche H, Eberle D. Automated compilation of pseudo-lithology maps from geophysical data sets: A comparison of Gustafson-Kessel and fuzzy c-means cluster algorithms[J]. *Exploration Geophysics*, 2011, 42(4):275–285.

[19] Deng C, Pan H, Luo M. Joint inversion of geochemical data and geophysical logs for lithology identification in CCSD main hole [J]. *Pure and Applied Geophysics*,2017, 174(12):4407–4420.

[20] Konaté A A, Ma H L, Pan H P, et al. Lithology and mineralogy recognition from geochemical logging tool data using multivariate statistical analysis[J]. *Applied Radiation and Isotopes*,2017,128: 55–67.

[21] Keykhay-Hosseinpppr M, Kohsary A-H. , Hossein-Morshedy A. , et al. A machine learning-based approach to exploration targeting of porphyry Cu-Au deposits in the Dehsalm district, Eastern Iran[J]. *Ore Geology Reviews*,2020,116(C):223–234.

[22] Rosenblatt M. Remarks on some nonparametric estimates of a density function [J]. *Ann. Math. Statist.*, 1956, 27(3): 832–837.

[23] E. Parzen. On estimation of a probability density function and mode[J]. *Ann. Math. Statist.*, 1962, 33(3): 1065–1076.

[24] Kazakos D. Choice of kernel function for density estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1980,2(3):255–258.

[25] Bolance C, Guillen M, Nielsen J P. Kernel density estimation of actuarial loss functions [J]. *Mathematics and Economics*, 2003, 32(1):19–36.

[26] Fukunaga K, Hostetler L D. The estimation of the gradient of a density function with applications in pattern recognition[J]. *Trans Information Theory*, 1975,21(2):32–40.

[27] 于传强, 郭晓松, 张安, 等. 基于估计点的滑动窗宽核密度估计算法[J]. *兵工学报*, 2009, 30(2):231–235.

Yu C Q, Guo X S, Zhang A, et al. Slide Bandwidth Kernel Density Estimation Algorithm Based on Estimate Point[J]. *Acta Armamentarii*, 2009,30(2):231–235.

Lithology identification based on Bayesian probability using adaptive kernel density

CAI Ze-Yuan^{1,2,3}, LU Bao-Liang^{1,2,3}, XIONG Sheng-Qing⁴, WANG Wan-Yin^{1,2,3}

(1. *Institute of Gravity and Magnetic Technology, Chang'an University, Xi'an 710054, China*; 2. *College of Geology Engineering and Geomatics, Chang'an University, Xi'an 710054, China*; 3. *Key Laboratory of Western China's Mineral Resources and Geological Engineering, Ministry of Education, Chang'an University, Xi'an 710054, China*; 4. *China Aero Geophysics and Remote Sensing Center for Natural and Resources, Beijing 100083, China*)

Abstract: Accurately characterizing rock types and their structural relationships can provide important information for researches on energy and mineral exploration, deep structure and tectonic, and so on. At present, geophysical data can be used to identify lithology through the differences between the corresponding physical property parameters of different rocks (such as density, susceptibility, resistivity, speed, etc.). However, different rock physical properties often coincide at a certain degree, on the other hand the result of lithological identification is not accurate enough when a single physical property is used. Therefore, it is of great significance to use multi-source data for lithological identification. Bayesian method belongs to the statistical classification methods, which relies on probability for classification, and the calculation of probability density depends on the correlation between sample attributes. On such a basis, the authors introduce the Bayesian probability model based on adaptive kernel density estimation to lithology identification. This method has good adaptability for many different types of physical property parameters and such the predicted lithology classification results with probability parameters, fuzzy intervals, and a variety of lithology classification results. This method has a strong scalability that can process both parametric and non-parametric information at the same time to maximally the known geological information and physical parameters. Synthetic models prove that this method has the ability to provide more stable and accurate results of lithology recognition compared with the methods of traditional Gaussian algorithm and fixed bandwidth kernel density estimation.

Key words: Bayesian; lithology identification; physical property parameter; adaptive kernel density