

# 强容噪性随机森林算法在地震储层预测中的应用

宋建国\*<sup>①②</sup> 杨璐<sup>①</sup> 高强山<sup>③</sup> 刘炯<sup>④</sup>

(<sup>①</sup>中国石油大学(华东)地球科学与技术学院,山东青岛 266580; <sup>②</sup>海洋矿产资源评价与探测技术功能实验室,山东青岛 266071; <sup>③</sup>中国科学院地球化学研究所月球与行星科学研究中心,贵州贵阳 550081; <sup>④</sup>中国石化石油勘探开发研究院,北京 100083)

宋建国,杨璐,高强山,刘炯. 强容噪性随机森林算法在地震储层预测中的应用. 石油地球物理勘探,2018,53(5): 954-960.

**摘要** 地震数据和测井数据中的噪声与有效信号难以有效界定,决定了地震储层预测需采用强容噪性算法。通过训练样本中加入随机噪声证实随机森林算法具有较好容噪性,但不能据此推知它在地震储层预测中仍有很强容噪性。基于 F3 工区实际数据,从噪声较强的原始地震数据中提取含噪样本,由经过倾角中值滤波处理的地震数据提取去噪样本,建立多种地震属性与孔隙度参数之间的随机森林回归模型;由构建的含噪模型和去噪模型分别与原始地震数据去噪前后两个数据体进行运算,得到 4 种不同情况下的孔隙度数据体。结果表明:由含噪模型得到的两个预测结果受噪声干扰较大;去噪模型的两个预测结果受噪声影响较小,能有效刻画储层特征,表现出强容噪性。随机森林模型对异于样本数据的异常值具有强的容忍度。可知随机森林算法应用于地震储层预测的关键是样本数据不含噪声,而估算过程中地震数据体是否做了去噪处理对预测结果影响较小。

**关键词** 地震属性 随机森林 容噪性 储层预测 噪声

中图分类号:P631 文献标识码:A doi: 10.13810/j.cnki.issn.1000-7210.2018.05.008

## 1 引言

综合地震、测井与地质资料进行解释与分析是油气勘探和开发过程中最基本、最重要的一种手段,也是油藏描述最基本的分析方法<sup>[1]</sup>。地震储层预测方法以地震属性数据为基础<sup>[2]</sup>,结合测井与地质信息进行储层特征参数预测<sup>[3]</sup>,可深入地研究特征参数的空间分布规律,减少地球物理反演问题的多解性,利用不同类型的资料优势,提高解释的可靠性<sup>[4-6]</sup>。

地震方法与测井方法在实际观测过程中受到各种干扰因素的影响,观测数据不可避免存在噪声,因而用于解释的数据不是完全的确定数据。地球物理各种反演方法都是建立在正演的基础上,对地下的

复杂地质问题做了“模型理想化”处理,对于实际数据中的有效信号与噪声信号的界限很难有一个明确的数学关系去直接界定<sup>[7,8]</sup>,滤波等去噪方法能够一定程度的消除地震数据中噪声,但无法完全消除噪声,使用一种对地震数据噪声具有较高容忍度的储层预测方法就显得非常重要。

随机森林是 Breiman<sup>[9,10]</sup>于 2001 年提出的一种组合预测算法,即通过组合大量决策树的预测结果作为一个整体的输出,有效地克服了单一决策树容易出现过拟合的问题。理论和实际数据的测试证明随机森林在很多领域对异常值和噪声具有很强的容忍度,克服了传统预测方法信息和知识的获取方式间接、费时且效率低下的缺点<sup>[11]</sup>。针对随机森林的容噪性,Dietterich<sup>[12]</sup>证明对训练样本的部分输出数

\* 山东省青岛市黄岛区长江西路 66 号中国石油大学(华东)地球科学与技术学院,266580。Email:jianguosong@163.com

本文于 2017 年 12 月 22 日收到,最终修改稿于 2018 年 7 月 13 日收到。

本研究受国家“十三·五”科技重大专项(2017ZX05049002)、国家自然科学基金项目(41674125)、国家自然科学基金委员会中国石油化工股份有限公司石油化工联合基金资助项目(U1663207)和国家重点基础研究发展计划项目(“973”计划项目)(2014CB239104)联合资助。

据加入随机噪声后, Bagging 方法和随机分割选择方法具有较强容噪性。Breiman 通过很多不同实际类型的数据对随机森林的容噪性进行了试验, 结果表明随机森林对噪声数据有很好的容忍度<sup>[13]</sup>。显然, 两位专家均是对输出数据加入随机噪声, 但所使用的多个输入变量是未加噪声的。其他学者在对随机森林的研究中, 通过实际数据和理论推导指出了输入变量在随机森林模型占有重要的地位, 并未对输入变量含有噪声时的情况进行分析<sup>[14-16]</sup>。

目前, 在与储层、岩性等预测相关的方面, 已有一些文献<sup>[17-20]</sup>对随机森林算法做了应用性的研究, 展现了随机森林分类或回归算法的优异性能, 相比神经网络算法, 随机森林算法操作简单, 运算效率较高<sup>[21]</sup>, 与支持向量机相比, 它需要设置的参数较少<sup>[22, 23]</sup>。但这些文献并未对输入变量或者属性数据含有噪声时对模型影响进行专门研究。地震数据和测井数据实质上均含有噪声, 即输入和输出变量均受到了噪声干扰, 随机森林对于输入数据, 即地震属性数据含有噪声时, 容噪性如何有待研究和分析。

基于以上的研究现状, 有针对性地分析随机森林算法在地震储层预测方面的容噪性显得很有必要。本文采用 F3 工区的实际数据研究随机森林回归算法(RRF)的容噪性。F3 工区原始地震数据含有较大噪声, 基于倾角导向滤波处理能够有效压制噪声, 得到高信噪比数据。从含噪声的原始数据和倾角导向滤波之后的高信噪比数据中分别提取样本并建立回归模型, 将这两个随机森林回归模型分别作用于原始数据和高信噪比数据, 预测储层特征参数, 用于对比分析。在建立随机森林回归模型和储层特征参数预测中, 提取了 8 种不同的地震属性作为随机森林回归算法的输入变量, 以孔隙度数据作为预测的特征参数, 即输出变量。两个随机森林回归模型分别作用于两个数据体, 计算得到 4 个孔隙度参数的数据体, 通过已知的地质信息分析这 4 个数据体之间的相关性, 发现样本是否含有噪声对回归模型的建立影响较大, 从而对预测结果的影响也较大。用高信噪比数据建立的回归模型对预测阶段数据中的噪声具有较好的容噪性。

## 2 随机森林回归算法原理

决策树是构成随机森林的基础, 在回归问题中,

决策树为回归树, 回归树的缺点主要在于容易出现过拟合, 只有少量的几个数值输出, 处理连续型变量问题能力局限性较大, 随机森林回归算法正是为解决这些问题而提出的。

随机森林回归算法利用 bootstrap 技术从原始样本集随机重复抽取多个样本子集, 对每个样本子集分别进行决策树建模, 预测时每棵决策树均给出一个预测结果, 将所有决策树输出值的平均作为森林预测值。随机森林基于统计学习理论, 融合了 bagging 方法和随机子空间方法两大机器学习算法。

回归森林的基本方法为:

(1) Bagging 思想: 首先, 利用 bootstrap 重复抽样方法从原始的训练样本中抽取  $k$  个样本子集, 且每个样本子集的容量与原始训练集的一样,  $k$  个样本子集分别建立  $k$  个回归树模型, 预测时对  $k$  个回归树输出值通过取平均得到回归森林输出值。该做法使每棵树的训练子集样本存在一定的重复, 目的是避免随机森林的决策树产生局部最优解。

(2) 随机子空间方法: 在生成每棵回归树时, 每个节点分裂不是考察所有可能的分割, 而是每个节点随机选取部分输入变量的可能分割并找出其中的最优分割进行分裂, 即随机特征选取方法。

通过重复抽样和随机特征选取两大随机方式, 其构建每棵回归树的样本和构建模型的过程均存在差异, 组合很多不同的回归树而形成森林, 从而提高了模型的外推预测能力。Breiman 通过理论和大量实际数据测试证明了随机森林算法不会轻易产生过拟合的问题, 随机森林的泛化误差小于决策树的平均泛化误差, 对数据中的噪声和异常值具有很好的容忍度, 对噪声具有很好的稳健性。

## 3 测试数据分析

分别采用模拟数据和实际数据分析, 说明随机森林回归算法对噪声具有很好的容忍度。

### 3.1 模拟数据试验

采用的模拟数据由以下非线性函数生成

$$y = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 - 7x_6 + 3x_7 - 2x_8 \quad (1)$$

式中  $x_1 \sim x_8$  是由计算机生成的具有高斯分布的模拟数据, 其样本点共计 400 个。对这 8 个变量分别加入信噪比为 10 : 1 的高斯白噪声, 其中一个变量

加入噪声前、后的对比情况如图 1 所示。

将变量  $x$  加噪前、后数据与  $y$  值分别组成样本集,从样本中选取  $3/4$  的数据作为训练样本, $1/4$  数据作为测试样本,利用训练样本分别建立不含噪与含噪模型,分别以加噪前、后的变量作为输入数据预测函数值,实际值与预测值对比情况如图 2 所示。

基于不含噪模型对不加噪数据和加噪数据进行预测,预测值与实际值的相关值分别为 91.07% 和 92.90%。含噪模型对不加噪数据和加噪数据的预测,预测值与实际值的相关值分别为 92.09% 和 91.23%。由以上分析可知,不含噪模型对噪声具有很好的容忍性。但是当随机森林模型的输入数据受

到噪声污染时,预测过程中即使经过了去噪处理,预测结果依然包含大量的噪声。

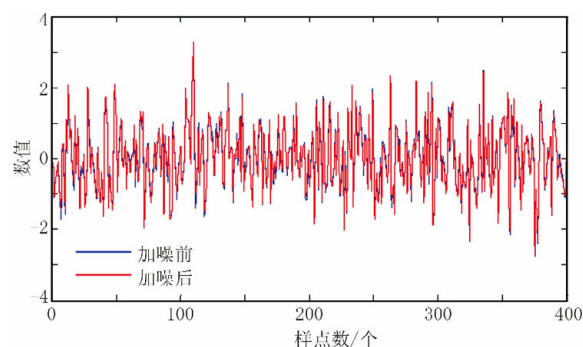


图 1 变量加噪前、后对比图

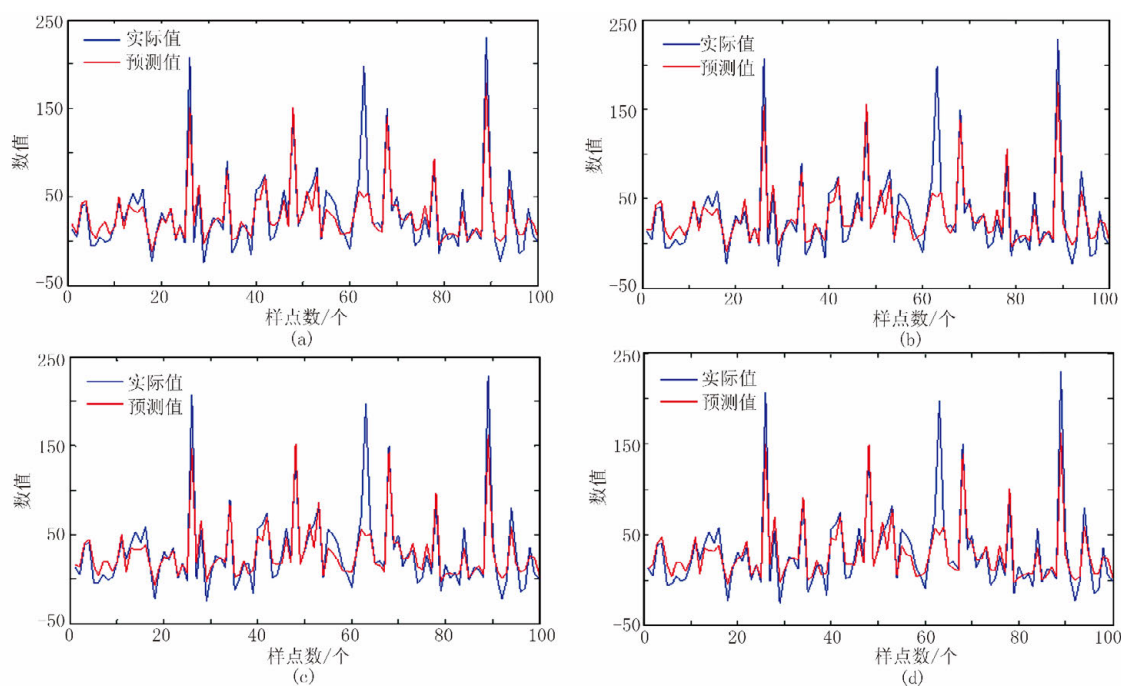


图 2 实际值与预测值对比图

(a) 不含噪模型对不含噪数据的预测; (b) 不含噪模型对含噪数据的预测;  
(c) 含噪模型对不加噪数据的预测; (d) 含噪模型对加噪数据的预测

### 3.2 实际数据试验

采用公开的国外海上 F3 工区实际数据研究随机森林在处理回归问题中的容噪性,以预测结果能否有效刻画储层的地质特征作为评价标准。

F3 是北海位于荷兰部分的一个区块,数据体上部 1200ms 内的反射层属于中新统、上新统和更新统。地震剖面上有一个非常明显的大型 S 型反射,是一个大型河控三角洲体系沉积体。工区内储层孔隙度总体很高(0.20~0.30),从地震剖面上可以看到典型的下超、顶超、上超和削截构造现象,生物成因气藏引起的亮点反射也清晰可见<sup>[24]</sup>。工区所有

井中的孔隙度数据均由密度计算得到

$$\text{孔隙度} = \frac{2.65 - \text{密度}}{2.65 - 1.05} \quad (2)$$

F3 工区演示数据体的原始地震数据体包含相当大的噪声,适合于分析当输入数据受到噪声干扰情况下,随机森林对噪声的容忍度或稳健性。

#### 3.2.1 地震数据滤波前、后信噪比分析比较

F3 演示数据体除了提供原始地震数据之外,还有由 Opentect 软件倾角控制模块进行中值滤波去噪处理的数据体,滤波去噪前后地震剖面如图 3 所示,可以看出经过滤波处理后,地震剖面上的同相

轴更加清晰,更有利于解释构造特征。在随后的试验分析中以原始地震数据作为含噪声数据,滤波处理后的地震数据作为去噪数据。倾角中值滤波是对地震数据做平滑处理,消除了其中的相关性较低的噪声,信噪比得到提高,但分辨率有一定程度的降低<sup>[25,26]</sup>。

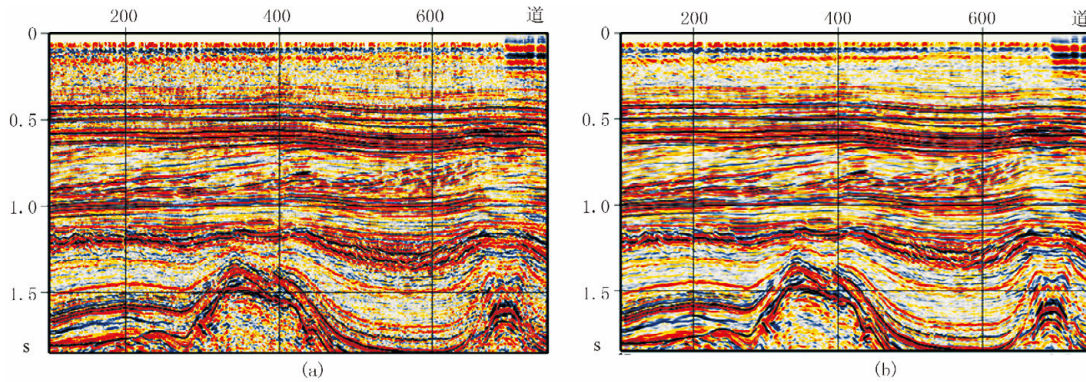


图 3 去噪前(a)、后(b)地震剖面

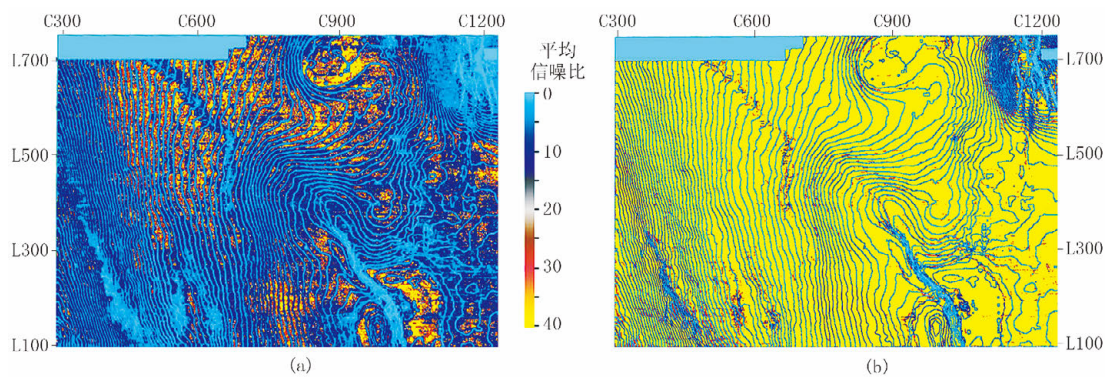


图 4 去噪前(a)、后(b)地震数据信噪比分析结果

### 3.2.2 试验方法及结果分析

将随机森林回归方法应用于地震储层预测,提取 F03-2 井上的孔隙度数据作为预测的储层特征参数,并与对应的井旁道地震属性组成 390 个原始样本,孔隙度参数的极值为 0.0568 和 0.3893,平均值为 0.3192。地震数据中的亮点反射比较明显,提取的地震属性以振幅类为主,包括 8 种常见的地震属性:平均绝对值振幅、平均振幅、均方根振幅、弧长、平均能量以及“三瞬”属性。将孔隙度数据与含噪声的地震属性数据和去噪后的地震属性数据分别组成两个样本集——含噪声样本集与不含噪声样本集。

从两个样本集中随机提取 3/4 作为训练样本,用于构建随机森林模型,剩下的 1/4 作为测试样本检验泛化性能,决策树设置为 500 棵,随机特征选取个数设置为 3,得到两个随机森林孔隙度预测模型,

图 4 为两个数据体在层位(Demo6)切片上的平均信噪比,由 SMI 地震波形特征指示反演软件分析获得。图中浅蓝色曲线为 SMI 软件分析信噪比的一系列控制点,可见滤波处理后大部分区域地震数据信噪比得到明显提高,个别区域资料信噪比改善微弱。

一个为含噪的随机森林模型,一个为不含噪的随机森林模型。将这两个模型应用于含噪数据(原始地震数据)和不含噪数据(中值滤波数据),就得到四个孔隙度预测结果。从含噪数据(原始地震数据)和不含噪数据(中值滤波数据)对应的层位(Demo6)切片中分别截取同一位置进行显示(图 4),从孔隙度数据体中提取的层位(Demo6)切片如图 5 所示。

图 6 展示的是由含噪声样本建立的 RF 模型,分别使用滤波前、后的地震数据所提取的地震属性为输入,估算得到的两个孔隙度数据。图 7 是不含噪 RF 模型应用于滤波前、后的地震数据所提取的地震属性,估算得到的两个孔隙度数据体。

比较预测结果可知,这四个预测结果在横向上所展示的孔隙度分布特点大致相似,均能有效地刻画在该层位上发育的南北向分布的河道特征<sup>[27,28]</sup>,

同时孔隙度数值主要为 0.25~0.35, 与实际的孔隙度(主要为 0.20~0.30)情况是符合的。结合图 4 中的信噪比, 可以看出在信噪比相对较高的区域, 四个预测结果所刻画的储层特征是基本相同的, 但在信噪比较低的区域, 预测结果所反映的地质信息存在较大差异, 由图 6 和图 7 可知导致出现差异的原因为所使用的模型样本数据是否含有噪声, 表明样本的模型训练过程中输入数据是否含有噪声对预测结果影响较大, 而在估算数据体的过程中地震属性数据中是否含有噪声对预测结果的干扰程度较小。

从地震剖面上分析, 如图 8 和图 9 所示, 可以看

出 RF 模型的样本数据是否含有噪声对预测结果影响较大, 图中黑色线为层位 Demo6, 该层位处于前积反射、超覆构造的上部。比较图 8 与图 9 的预测结果可知, 当地震属性样本数据不含有噪声时, 预测结果更加符合实际情况, 对 F3 工区的超覆现象刻画效果更好, 去噪模型的预测结果(图 9)分辨率有所降低, 其原因在于所使用的滤波去噪方法为倾角中值滤波, 提高信噪比的同时付出了降低分辨率的代价。

试验结果表明输入数据在随机森林建模中起很重要的作用, 训练样本的输入数据受到噪声污染时,

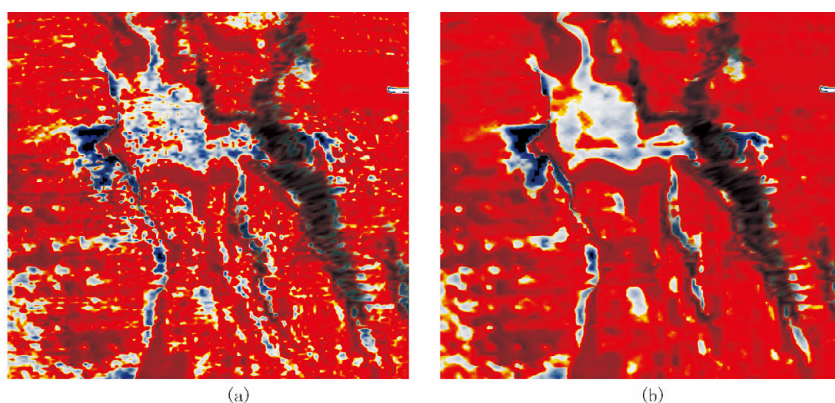


图 5 去噪前(a)、后(b)地震数据层位切片

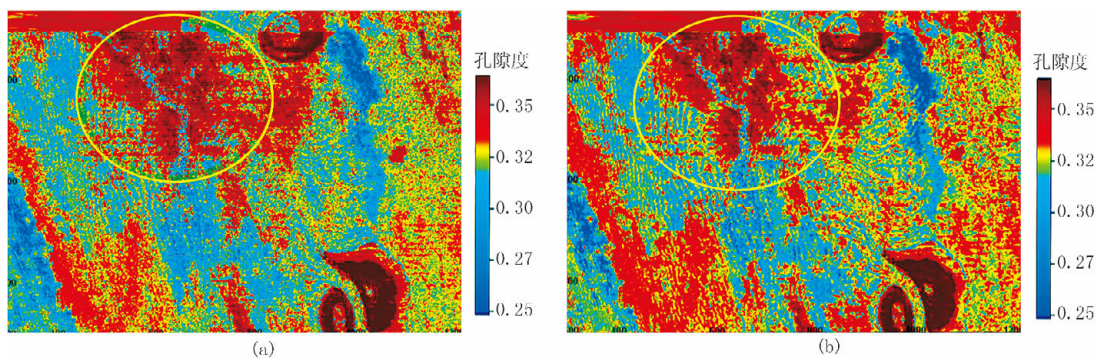


图 6 含噪模型对地震数据滤波前(a)、后(b)预测的孔隙度层位切片

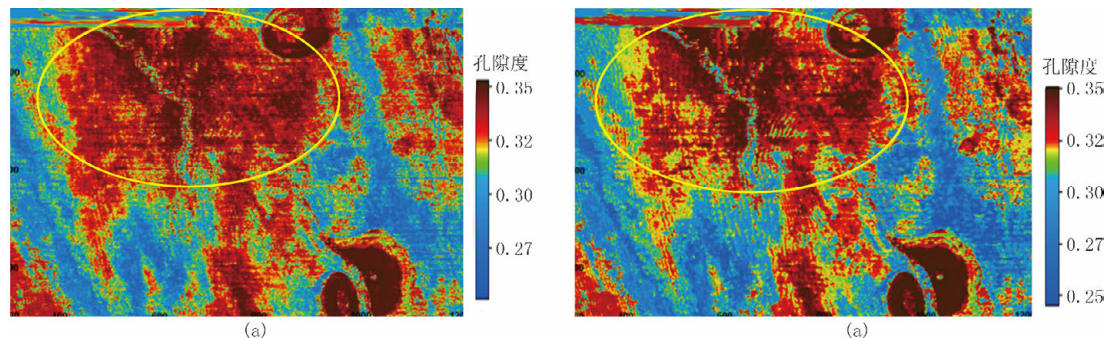


图 7 不含噪模型对地震数据滤波前(a)、后(b)预测的孔隙度层位切片

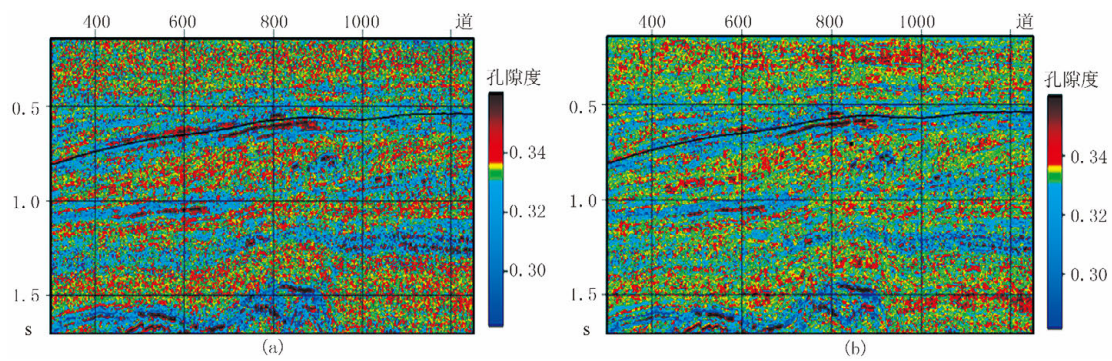


图 8 含噪模型对地震数据滤波前(a)、后(b)预测的孔隙度剖面

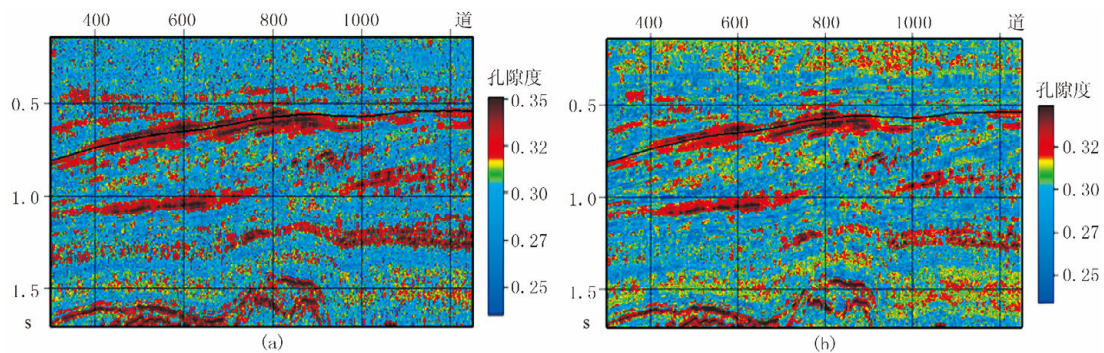


图 9 去噪模型对地震数据滤波前(a)、后(b)预测的孔隙度剖面

输出结果受到噪声的影响较大,而在模型建立之后,估算数据体时地震属性数据是否做了去噪处理对预测结果的影响较小,可知当对样本的输入数据做了滤波处理,即在建立的模型受到噪声污染较少情况下,能表现出很好的容噪性。

## 4 讨论与认识

通过 F3 工区实际数据的容噪性试验分析,表明当训练样本的输入数据不含噪时,随机森林回归算法具有很强的容噪性,在预测过程中随机森林模型受噪声干扰较小。当训练样本的输入数据受噪声污染时,即使在预测过程中进行去噪处理,地震储层预测结果依然可能包含大量噪声。容噪性试验结果同时证明了随机森林算法具有较强稳健性,用于构建随机森林模型的样本数据在储层预测中占据很重要地位;在模型建立之后进行预测时,随机森林模型对异于样本信息的数据具有较强的容忍度。因此采用随机森林算法进行地震储层预测时,关键是提取不含噪声的样本数据。

### 参考文献

- [1] 王永刚.地震资料综合解释方法.山东东营:中国石油大学出版社,2007.
- [2] 李伟,岳大力,胡光义等.分段地震属性优选及砂体预测方法——秦皇岛 32-6 油田北区实例.石油地球物理勘探,2017,52(1):121-130.  
Li Wei, Yue Dali, Hu Guangyi et al. Frequency segmented seismic attribute optimization and sandbody distribution prediction; an examples in North Block, Qinhuangdao 32-6 Oilfield. OGP, 2017, 52(1):121-130.
- [3] 赵学松,高强山,唐传章等.基于支持向量回归机与并导向的三角洲岩性油气藏储层参数预测.石油地球物理勘探,2016,51(5):976-982.  
Zhao Xuesong, Gao Qiangshan, Tang Chuazhang et al. Delta stratigraphic reservoir parameter estimation based on support vector regression machine and well logging data. OGP, 2016, 51(5):976-982.
- [4] 乐友喜,杨丽.储层地震预测基础理论方法研究.天然气地球科学,2009,20(4):563-570.  
Yue Youxi, Yang Li. Basic theories and methods of seismic reservoir prediction. Natural Gas Geosciences, 2009, 20(4):563-570.
- [5] 印兴耀,周静毅.地震属性优化方法综述.石油地球物理勘探,2005,40(4):482-489.  
Yin Xingyao, Zhou Jingyi. Summary of optimum methods of seismic attributes. OGP, 2005, 40(4):482-489.
- [6] 赵小辉,于宝利,曹小璐.属性融合技术在微小断裂识别中的应用.石油地球物理勘探,2017,52(增刊 2):164-169.  
Zhao Xiaohui, Yu Baoli and Cao Xiaolu. Minor fault identification with seismic multi-attribute fusion. OGP, 2017, 52(S2):164-169.

- [7] 杨平,孙赞东,许辉群等.叠后储层地震滤波质量控制方法.地球物理学进展,2014,29(1):234-239.  
Yang Ping, Sun Zandong, Xu Huiqun et al. Quality control of reservoir seismology filtering method. Progress in Geophysics, 2014, 29(1): 234-239.
- [8] 陆基孟,王永刚.地震勘探原理.山东东营:中国石油大学出版社,2009.
- [9] Breiman L. Random forests. Machine Learning, 2001, 45(1):5-32.
- [10] Breiman L. Consistency for a Simple Model of Random Forests. Technical Report 670, UC Berkeley, CA, 2004, URL www.stat.berkeley.edu/breiman/.
- [11] 方匡南.随机森林组合预测理论及其在金融中的应用.福建厦门:厦门大学出版社,2012.
- [12] Dietterich T. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization. Machine Learning, 2000, 40(2):139-157.
- [13] Genuer R, Poggi J M, Tuleau C. Random Forests: Some Methodological Insights. Inria 6729, 2008, 32, arXiv:0811.3619 v1[stat. ML], ISSN0249-6399.
- [14] Biau G. Analysis of a random forests model. Journal of Machine Learning Research, 2012, 13(2):1063-1095.
- [15] Prasad A M, Iverson L R, Liaw A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems, 2006, 9(2):181-199.
- [16] 宋建国, 高强山, 李哲. 随机森林回归在地震储层预测中的应用. 石油地球物理勘探, 2016, 51(6):1202-1211.  
Song Jianguo, Gao Qiangshan and Li Zhe. Application of random forests for regression to seismic reservoir prediction. OGP, 2016, 51(6):1202-1211
- [17] 王志宏, 韩璐, 戚磊. 随机森林分类方法在储层岩性识别中的应用. 辽宁工程大学学报:自然科学版, 2015, 34(9):1083-1088.  
Wang Zhihong, Han Lu, Qi Lei. Random forests classification method in the application of reservoir lithology recognition. Journal of Liaoning Technical University: Natural Science, 2015, 34(9):1083-1088.
- [18] 赵保成, 谭志详, 邓喀中. 利用随机森林回归模型计算主要影响角正切. 金属矿山, 2016, 45(3):172-175.  
Zhao Baocheng, Tan Zhixiang, Deng Kazhong. Calculation of the tangent of major influence angle based on random forest regression model. Metal Mine, 2016, 45(3):172-175.
- [19] 温廷新, 张波, 邵良杉. 煤与瓦斯突出预测的随机森林模型. 计算机工程与应用, 2014, 50(10):233-237.  
Wen Tingxin, Zhang Bo, Shao Liangshan. Prediction of coal and gas outburst based on random forest model. Computer Engineering and Application, 2014, 50(10):233-237.
- [20] Kuhn S, Cracknell M J, Reading A M. Lithological mapping via random forests, Information entropy as a proxy for inaccuracy. ASEG Extended Abstracts 2016; 25th International Geophysical Conference and Exhibition, 904-907.
- [21] 程国建, 周冠武, 王潇潇. 概率神经网络方法在岩性识别中的应用. 微计算机信息, 2007, 23(6-1):288-289.  
Cheng Guojian, Zhou Guanwu, Wang Xiaoxiao. The probability neural networks for lithology identification. Microcomputer Information, 2007, 23(6-1):288-289.
- [22] 宋延杰, 张剑风, 闫伟林等. 基于支持向量机的复杂岩性测井识别方法. 大庆石油学院报, 2007, 31(5):18-21.  
Song Yanjie, Zhang Jianfeng, Yan Weilin et al. A new identification method for complex lithology with support vector machine. Journal of Daqing Petroleum Institute, 2007, 31(5):18-21.
- [23] 张长开, 姜秀娣, 朱振宇等. 基于支持向量机的属性优选和储层预测. 石油地球物理勘探, 2012, 47(2):282-285.  
Zhang Changkai, Jiang Xiudi, Zhu Zhenyu et al. Attributes selection and reservoir prediction based on support vector machine. OGP, 2012, 47(2):282-285.
- [24] http://www.opendtect.org/osr/Main/Netherlands-OffshoreF3BlockComplete4GB.
- [25] Tingdahl K M, De Groot P F M. Post-stack dip- and azimuth processing. Journal of Seismic Exploration, 2003, 12(2):113-126.
- [26] 赵丽娅, 傅群, 潘海滨. Opendtect 软件倾角控制模块在地震解释中的应用. 海洋地质动态, 2008, 24(4):33-37.  
Zhao Liya, Fu Qun, Pan Haibin. Application of Opendtect software dip-steering to seismic interpretation. Marine Geology Letters, 2008, 24(4):33-37.
- [27] Overeem I, Weltje G J, Bishop-Kay C et al. The Late Cenozoic Eridanos delta system in the Southern North Sea Basin, a climate signal in sediment supply?. Basin Research, 2001, 13(3):293-312.
- [28] Sørensen J C, Gregersen U, Breiner M et al. High frequency sequence stratigraphy of upper Cenozoic deposits. Marine Petroleum Geology, 1997, 14(2):99-123.

(本文编辑:朱汉东)

## 作者简介



宋建国 副教授, 硕士生导师, 1971年生; 1993年毕业于中国石油大学(华东)物探专业, 获学士学位; 分别于1996、1999年获中国石油大学(华东)应用地球物理专业硕士和博士学位; 现在中国石油大学(华东)地球科学与技术学院从事地震资料高分辨率处理与解释理论、方法的教研。

greater than 97% under the noise free condition. Tests with different noise levels indicate that the proposed method is more stable than the amplitude ratio method and can improve the inversion accuracy. Finally, real data tests in an oil field show that the proposed method achieves good results in application.

**Keywords:** downhole microseismic monitoring system, focal mechanism, wave energy, grid search

1. Institute of Oil & Gas, School of Earth and Space Sciences, Peking University, Beijing 100871, China

2. Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

3. Shengli Branch, Geophysical Company, SINOPEC, Dongying, Shandong 257086, China

**Strong tolerance random forest algorithm in seismic reservoir prediction. Song Jianguo<sup>1,2</sup>, Yang Lu<sup>1</sup>, Gao Qiangshan<sup>3</sup> and Liu Jong<sup>4</sup>. OGP, 2018, 53(5): 954-960.**

Since it is difficult to effectively define noise and signal on seismic and logging data, seismic reservoir prediction needs good noise tolerance algorithms. The random forest (RF) algorithm with strong noise tolerance is proved by adding noise to training samples. However this does not mean RF has good noise tolerance in seismic reservoir prediction as well. First we extract noise samples from original seismic data with strong noise in the Survey F3, and extract denoised samples from the data processed by the dip-steered median filter. Then we establish random forest regression models between seismic attributes and the porosity parameter. After processing the original seismic data and the filtered seismic data with the noise sample model and denoised sample model, we estimate four different porosity parameter cubic data. The results reveal that the two data sets obtained with the noise model are more disturbed by noise, and the other two data sets obtained with the denoised model are much less affected by noise. On these data sets, reservoir geological characteristics can be effectively characterized which proves the random forest model has strong robustness and perfect tolerance to abnormal data differing from the sample data. The key issue in the application of the random forest algorithm to seismic reservoir prediction is that the training data does not contain noise. In other words, the input variable of sample data being denoised is much more significant, whereas whether seismic data were denoised or not has less effects on the prediction result.

**Keywords:** seismic attribute, random forest algorithm, noise tolerance, reservoir prediction, noise

1. School of Geosciences, China University of Petroleum (East China), Qingdao, Shandong 266580, China

2. Laboratory of Marine Mineral Resource Evaluation and Detection, Qingdao National Laboratory for Marine Science and Technology, Qingdao,

Shandong 266071, China

3. Lunar and Planetary Science Research Center, Institute of Geochemistry, Chinese Academy, Guiyang, Guizhou 550081, China

4. Research Institute of Petroleum Exploration and Production, SINOPEC, Beijing 100083, China

**Seismic sparse spike inversion based on  $L_0$  norm approximation. Liu Baihong<sup>1</sup>, Li Jianhua<sup>2</sup> and Zheng Silian<sup>1</sup>. OGP, 2018, 53(5): 961-968.**

Under the assumption that the reflectivity is a series of sparse spike, a sparse spike inversion for seismic reflection coefficients is proposed. First the inversion is accomplished by minimizing an objective function which includes a quadratic error term ( $L_2$  norm) combined with a sparseness-inducing regularization term ( $L_0$  norm). Then the  $L_0$  norm is approximated by a smoothing function, so that the  $L_0$  penalty becomes differentiable, and the minimization algorithm for continuous functions based on derivative is used to solve the problems. The proposed method is tested on both model data and real field data, and the results show this method is feasible.

**Keywords:** seismic exploration, reflectivity, impedance, sparse spike inversion,  $L_0$  norm

1. Geophysical Research Institute, SINOPEC, Nanjing, Jiangsu 211103, China

2. Geological Research Center, GRI, BGP Inc., CNPC, Zhuozhou, Hebei 072750, China

**AVO types discrimination based on a proximal support vector machine. Li Wenxiu<sup>1,2,3</sup>, Wen Xiaotao<sup>1,2</sup>, Li Tian<sup>3</sup>, Li Leihao<sup>1</sup>, Liu Songming<sup>1</sup> and Yang Jixin<sup>1</sup>. OGP, 2018, 53(5): 969-974.**

AVO is an important approach for reservoir oil and gas analysis. It can qualitatively describe oil reservoirs. The AVO conventional classification depends mainly on human discrimination so that the discrimination result is often inaccurate and the workload is heavy. In this paper, we extract feature parameters from four types of AVO curves as a training set, and introduce the proximal support vector machine method to AVO types discrimination. Based on the shape of four types of gas AVO curves, taking the morphological features of pre-stack seismic data as input parameters, AVO types of the reservoir in a survey area are obtained. This method is applied to the automatic identification of AVO types in a clastic-rock gas field in the South China Sea, and more accurate results are obtained. The proposed method provides a reliable and convenient tool for AVO types discrimination in reservoirs.

**Keywords:** proximal support vector machine, AVO types, classification, reservoir analysis

1. Institute of Geophysics, Chengdu University of Technology, Chengdu, Sichuan 610059, China

2. State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu, Sichuan 610059, China