Journal of Computer Applications
计算机应用,2021,41(8):2453－2459
ISSN 1001-9081
CODEN JYIIDU
2021-08-10
http://www.joca.cn

文章编号:1001-9081(2021)08-2453-07
DOI:10.11772/j.issn.1001-9081.2020101867

基于多粒度时序结构表示的异常检测算法在储层含油性检测中应用

孟　凡[1],陈　广[1],王　勇[2],高　阳[1*],高德群[2],贾文龙[3]
(1.南京大学 计算机科学与技术系,南京 210023;　2.中国石油化工股份有限公司江苏油田分公司 物探技术研究院,南京 210046;
3.南京天技通信技术实业有限公司,南京 210019)
(*通信作者电子邮箱 gaoy@nju.edu.cn)

摘　要:传统储层含油性勘测方法利用地震波穿过地层时产生的相关地震属性和地质钻井资料结合传统地球物理方法进行综合研判,但该类勘测方法往往存在研判成本高且对专家先验知识依赖性强的问题。针对该问题,以江苏油田苏北盆地的地震资料为基础,并结合含油样本的稀疏性和随机性,提出了一种基于多粒度时序结构表示的异常检测算法,直接利用叠后地震道数据进行预测。该算法首先对于单个地震道数据提取多粒度时序结构并形成独立特征表示;其次,在提取多个粒度时序结构表示的基础上进行特征融合,以形成对地震道数据的融合表示;最后,通过对融合后的特征采用代价敏感方法进行联合训练和判别,从而得到对于该地震数据的含油性勘测结果。所提算法在江苏油田实际原始地震资料上进行了实验仿真,实验结果表明:所提算法相比长短期记忆(LSTM)和门控循环单元(GRU)算法在曲线下方的面积(AUC)指标上均提升了10%。

关键词:含油性检测;异常检测;多粒度;代价敏感;时序结构
中图分类号:TP391　　**文献标志码**:A

Multi-granularity temporal structure representation based outlier detection method for prediction of oil reservoir

MENG Fan[1], CHEN Guang[1], WANG Yong[2], GAO Yang[1*], GAO Dequn[2], JIA Wenlong[3]
(1. *Department of Computer Science and Technology, Nanjing University, Nanjing Jiangsu* 210023, *China*;
2. *Geophysical Prospecting Research Institute, Jiangsu Oilfield Company, Sinopec Group, Nanjing Jiangsu* 210046, *China*;
3. *Nanjing Tach Communication Technology Industry Company Limited, Nanjing Jiangsu* 210019, *China*)

Abstract: The traditional methods for prediction of oil reservoir utilize the seismic attributes generated when seismic waves passing through the stratum and geologic drilling data to make a comprehensive judgment in combination with the traditional geophysical methods. However, this type of prediction methods has high cost of research and judgement and its accuracy strongly depends on the prior knowledge of the experts. To address the above issues, based on the seismic data of the Subei Basin of Jiangsu Oilfield, and considering the sparseness and randomness of oil-labeling samples, a multi-granularity temporal structure representation based outlier detection algorithm was proposed to perform the prediction by using the post-stack seismic trace data. Firstly, the multi-granularity temporal structures for the single seismic trace data was extracted, and the independent feature representations were formed. Secondly, based on extracting multiple granularity temporal structure representations, feature fusion was carried out to form the fusion representation of seismic trace data. Finally, a cost-sensitive method was utilized for the joint training and judgement to the fused features, so as to obtain the results of oil reservoir prediction for these seismic data. Experiments and simulations of the proposed algorithm were performed on an actual seismic data of Jiangsu Oilfield. Experimental results show that the proposed algorithm is improved by 10% on Area Under Curve (AUC) compared to both of the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) algorithms.

Key words: prediction of oil reservoir; outlier detection; multi-granularity; cost-sensitive; temporal structure

0　引言

石油作为当今世界上举足轻重的战略物资,与国家的工业、农业、交通、国防以及每个人的衣食住行息息相关。然而,近年来江苏油田面对复杂严峻的生产经营形势,对岩层含油性勘测工作提出了更高的要求。传统岩层含油性检测任务是一项具有很大挑战的勘探工作,主要受到以下三个方面制约:

收稿日期:2020-10-29;**修回日期**:2020-12-30;**录用日期**:2020-12-30。

作者简介:孟凡(1988—),男,江苏南京人,高级工程师,博士研究生,CCF会员,主要研究方向:机器学习、无监督异常检测;　陈广(1992—),男,江苏南京人,硕士研究生,主要研究方向:数据挖掘;　王勇(1962—),男,河北吴桥人,教授级高级工程师,博士,主要研究方向:地震资料数据处理;　高阳(1972—),男,江苏淮安人,教授,博士生导师,博士,CCF高级会员,主要研究方向:大数据分析、机器学习、多智能体系统、视频/图像处理;　高德群(1970—),男,江苏泰州人,高级工程师,硕士,主要研究方向:石油地质勘探;　贾文龙(1984—),男,陕西商洛人,工程师,主要研究方向:大数据、人工智能。

www.joca.cn

1）领域性强。传统含油性检测方法需引入较强的专家先验知识进行综合研判，增加了实际勘探过程的不可复制性。即在实际含油性勘探工作中，通过传统地球物理方法得到初步结果后，往往需要再进行多轮专家综合研判，方可进行后续工程实施计划的制定。

2）验证成本高。通过钻井作业对给定工区含油性进行验证经济开销较大，且因误判造成的损失高昂。在实际的含油性勘探任务中，对地下岩层含油储层的误判将产生高昂的勘测代价，这也是制约现有勘测工作的重要因素之一。因此以原始地震资料为基础，结合少量测井标记信息对储层含油性预测至今仍是油气勘探开发与研究的热点[1]。简言之，探索一种能够利用少量现有测井标记信息且可降低假阳性和假阴性出现情况的方法将有着十分重要的社会价值和经济效益。

3）预测精度不理想。地下岩层存在着复杂的结构关系，给传统含油性勘测方法的预测精度带来挑战。传统地层含油性勘测方法为利用地震波穿过地层时产生的相关地震属性和地质钻井资料结合传统地球物理方法进行研判。但从实际检测效果来看，精度仍有进一步提高空间。

本文以实际含测井标记的叠后地震道数据为研究对象，通过结合物探专业领域知识和机器学习方法，从以下 3 个方面对含油性检测任务开展研究：

1）数据驱动学习。基于传统地球物理学含油性检测方法中涉及了大量专家领域知识进行综合研判，给实际勘探工作增加了不可复制性。因此，所设计算法需在一定程度上实现自动化检测，尽量减少对专家领域知识的依赖性（针对制约 1 进行探索）。

2）监督信息少且类别不平衡。由于含油层通常较薄，本身厚度通常仅有 1~4 m，受目前地震波采样设备频率限制，已钻井验证后的地震道样本与总体待检测样本相比极度稀少且类别不平衡。甚至油层在地震道数据样本中往往仅体现为单个独立点。因此所设计算法需针对极少量且类别不平衡的监督信息进行挖掘和研究，充分利用现有钻井样本集中的监督信息（针对制约 2 进行探索）。

3）数据呈时序性。叠后地震道数据是通过地震波为介质进行采集，由于其物理产生机制原因，该样本内数值之间是具有时序特性的地震属性值。因此所设计算法需挖掘出其内部潜在时序结构，并将此结构融入后续的检测过程中，进而提高检测精度（针对制约 3 进行探索）。

基于以上 3 个研究动机，本文从含油性储层检测任务中呈现的实际问题出发，提出了一种基于多粒度时序结构表示的异常检测算法，并利用江苏油田实际叠后地震资料进行仿真实验，本文主要包含以下 3 项工作：

1）针对地下储层含油性样本等监督信息稀少且类别不平衡现状，本文将该检测问题建模为有监督异常检测任务。即将含油样本视为异常样本进行检测。

2）提出了一种基于多粒度时序结构表示的异常检测算法。该算法首先针对单地震道数据中存在的潜在时序结构，提出基于双向长短期记忆（Bi-directional Long Short-Term Memory，BiLSTM）模型[2]的多粒度特征提取方法，捕捉了地震道内不同粒度的时序结构；其次，将不同粒度的时序结构及其特征进行自动化融合，进而形成了对单个地震道数据的多粒度时序融合特征；最后，针对该融合后的特征，利用代价敏感损失进行含油性异常样本的预测。

3）在江苏油田真实地震资料上进行了实验，实验结果从 4 个方面验证了本文所提出算法的有效性。

1 相关工作

近年来，随着机器学习在视觉和语音数据等方面取得了突破性进展[3-7]，石油勘探领域研究团队逐渐开始研究和探索将机器学习技术应用到储层含油性检测任务中。相比于传统含油性检测技术，即联合几何地震学和物理地震学特征分析方法对地震目标进行处理、地震属性提取优化与分析，近年来国内外研究人员不断尝试将机器学习方法和含油性检测相结合。该类方法利用有限的已知地质测井资料在叠后地震资料上构建训练样本，借助机器学习方法提取其中的潜在特征并建模。其中主要包括两大类方法：基于浅层学习的含油性检测方法[8-25]和基于深度学习的含油性检测方法[1,26-30]。本文主要聚焦于基于深度学习的含油性检测方法。

基于浅层学习方法在储层含油性检测场景中进行应用主要是利用机器学习技术中经典的浅层模型在地震道数据中进行训练和识别，其中涉及聚类分析[8]、人工神经网络[9-12,17-21]、模糊理论[13]、决策树[14]、遗传算法[15-16]、支持向量机[22]和概率模型等方法[23-25]。例如刘瑞林等[17]从监督学习角度，梳理了利用测井密度和自然电位等测井数据作为监督信息，用浅层神经网路算法对含油性特征进行分类等方法。陆文凯等[18]将测井资料作为监督学习样本，予以推广全区，进而获取横向上的含油性预测分类结果。张奕等[19]结合自组织神经网络算法和粒子群算法，获得地震相的聚类结果，该方法通过降低原始属性维度，提高了计算效率。张向君等[20]构建了基于结构风险最小化的神经网络，对吉林油田的储层油性特征进行了预测。Mobebbi 等[21]直接利用人工神经网络对测井数据进行预测。张长开等[22]提出了基于支持向量机的特征选择方法对地震属性优选并在此基础上预测储层。王波等[23]提出了基于斑块饱和模型井控属性融合法进行含油性检测。袁照威等[24]通过分析地震属性与沉积相的关系，运用马尔可夫-贝叶斯模拟算法，对多地震属性进行融合，得到多变量融合概率场信息，进而建立沉积相模型。宋建国等[25]以多种不同的地震属性为基础，通过构建井旁道地震属性与特征参数的回归森林模型进行储层预测。

以上基于浅层学习的方法在一定程度上缓解了对于专家领域知识的依赖。但由于储层含油性识别问题实际需要求解测井数据与实际地质中各组成部分之间的复杂映射关系，加之此种映射通常被认为是高度非线性问题，因此浅层学习模型的复杂度可能成为潜在制约。而深度方法因其无需建立显性复杂函数即可实现复杂非线性映射关系在近些年被进一步探索。基于深度学习方法在含油性检测技术中进行应用主要为利用深度卷积网络可隐式地构建地震属性与含油性储层分布的复杂映射关系。例如 Saikia 等[26]对浅层和深度神经网络提取储层含油性特征进行了综述。Gholami 等[1]以原始地震数据为基础获得地震属性进行油性储层预测。林年添等[27-28]通过引入卷积神经网络，智能提取、分类并识别地震资料中含油性特征，同时利用不同卷积核卷积形成各类纵、横波地震属性，进而对含油性储层地震预测。付超等[29]提出了一种卷积神经网络与支持向量机方法相结合的多波地震油气储层分布预测的深度学习方法。安鹏等[30]利用长短期记忆（Long Short-Term Memory，LSTM）模型捕捉测井参数和泥质含量（孔隙度）之间非线性映射关系进行储层物性参数预测。

相对于浅层学习方法，虽将深度学习直接应用于地震资

www.joca.cn

料处理领域取得了一定的成效,但该类方法依然面临着一些挑战:一方面由于地震道数据是地震波从激发点出发,经过地层滤波器返回到接收点的过程记录,因而其本身存在较为复杂的时序等结构关系,若忽略地震道数据之间结构关系而直接将其应用在深度网络中,可能将会对检测精度产生负面影响;另一方面由于井段验证样本稀少,有油与无油样本存在类别不平衡现象,为了防止深度模型陷入过学习或欠学习状态,现有方法往往过多依赖于人工经验对于模型超参数进行调测,若设置不当会使能实际检测性能产生较大波动。

针对以上分析,考虑到地震资料潜在时序结构和类别不平衡特性,本文提出了将储层含油性检测问题建模为面向时序数据的有监督异常检测任务,将含油性地震道样本作为异常点进行捕获。同时,考虑到单个地震道数据包含振幅采样点多和分布不均等特点,若直接在原始时间序列上进行含油性检测可能无法有效表达出其特征,进而出现判别精度不高现象。因此本文提出了一种基于多粒度时序结构表示的异常检测方法。

2　问题分析

本章主要介绍含测井标记地震道数据符号表示以及对于含油性检测任务形式化描述。

2.1　含测井标记地震道数据

含测井标记地震道数据是指利用原始叠后地震道数据对地下岩层进行分析后,再通过实际测井工作对该区域岩性进行实际验证后,所得到含有岩性标记的地震资料。

给定一个工区的原始地震道资料和地质测井资料,该区域的含测井标记地震道数据集合表示为 $\boldsymbol{D}=\{(\boldsymbol{x}_1,\boldsymbol{y}_1),(\boldsymbol{x}_2,\boldsymbol{y}_2),\cdots,(\boldsymbol{x}_j,\boldsymbol{y}_j)\}$,其中 $\boldsymbol{x}_i=\{x_i^{(1)},x_i^{(2)},\cdots,x_i^{(t)}\}$ 表示单个地震道数据,$x_i^{(t)}\in\mathbf{R}$ 表示在第 i 个地震道数据中第 t 时刻的地震属性值。$\boldsymbol{y}_i=\{y_i^{(1)},y_i^{(2)},\cdots,y_i^{(t)}\}$ 表示给定地震道数据内单个时间点对应含油性(岩层性质)标记,其中 $y_i^{(t)}\in\{0,1\}$。本文对单个地震道在某一时刻对应的地震属性采用一维原始振幅数据作为输入,若勘探领域业务人员后续需增加其他地震属性作为输入数据(例如:频率、瞬时相位等属性),可将以上形式化表示直接进行维度扩展。图 1 为江苏油田工区经过数据处理后的实际地震剖面图。

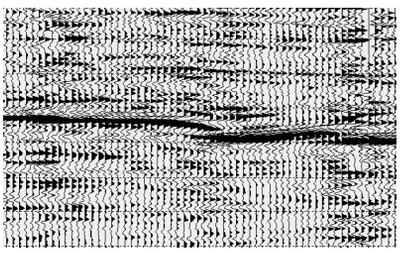

图 1　原始地震道数据片段

Fig. 1　Fragment of raw seismic trace data

2.2　含油性检测问题形式化描述

对于给定的未知地震道数据含油性检测问题,本文将其建模为有监督异常检测任务,可进一步形式化为:当给定一组已知含测井标记地震道数据 $\boldsymbol{D}$,含油性异常检测任务是通过学习目标函数(1),对新给定的地震道数据 $\hat{X}_i$ 进行含油性判别,即判别该地震道含油性概率。

$$\min_{\theta_\Psi,\theta_F,\theta_L,\theta_\Omega}\sum_{i=1}^{n}L(F(\Psi(\boldsymbol{x}_i|\theta_\Psi)|\theta_F),\boldsymbol{y}_i|\theta_L)+\Omega(\theta_\Omega)\tag{1}$$

其中:$\Psi(\cdot)=\sum_{i=1}^{n}\theta_{\Psi_i}$。$\boldsymbol{h}_i$ 为多粒度特征抽取与融合函数,表示从原始数据中提取不同粒度时序结构特征,并将其进行融合;θ_{Ψ_i} 表示第 i 个粒度特征的模型参数,$\boldsymbol{h}_i$ 表示第 i 个粒度的特征表示;$F(\cdot)=f_D(\boldsymbol{H}_i|\theta_F)$ 为含油性判别函数,表示将融合后特征映射为预测结果,$f_D(\cdot)$ 为全连接神经网络分类器,$\boldsymbol{H}_i$ 是融合后的特征,$\theta_F=(\boldsymbol{W}_D,\boldsymbol{b}_D)$ 为模型参数。$L(\boldsymbol{y}_i,\hat{\boldsymbol{y}}_i|\theta_t)$ 为损失函数,表示所提出模型预测结果与真实测井标记的损失代价,$\boldsymbol{y}_i$ 表示对含油性的真实标记,$\hat{\boldsymbol{y}}_i$ 为模型预测后的结果,$\theta_L=(\boldsymbol{W}_L,\boldsymbol{b}_L)$ 为模型参数。$\Omega(\cdot)$ 为正则化项,表示对提出模型的正则项约束。$(\theta_\Psi,\theta_F,\theta_L,\theta_\Omega)$ 为以上目标函数中所涉及的待优化模型参数。

3　算法设计

本文针对原始地震道数据中含油性检测问题,提出了一种基于多粒度时序结构表示的异常检测算法。该方法主要包括以下两个部分:1)多粒度时序结构特征抽取模块;2)基于多粒度时序结构特征融合与判别模块。

本文算法总体流程如图 2 所示:首先,多粒度时序结构提取模块在叠后地震资料中取出单个地震道数据,并对该道数据进行长序列划分。针对得到的长序列数据进行趋势性结构、周期性结构和局部性结构提取并形成多个粒度的独立特征。其次,在特征融合与判别模块将针对多个已取得的独立特征进行融合,进而在融合特征基础上进行异常判别。

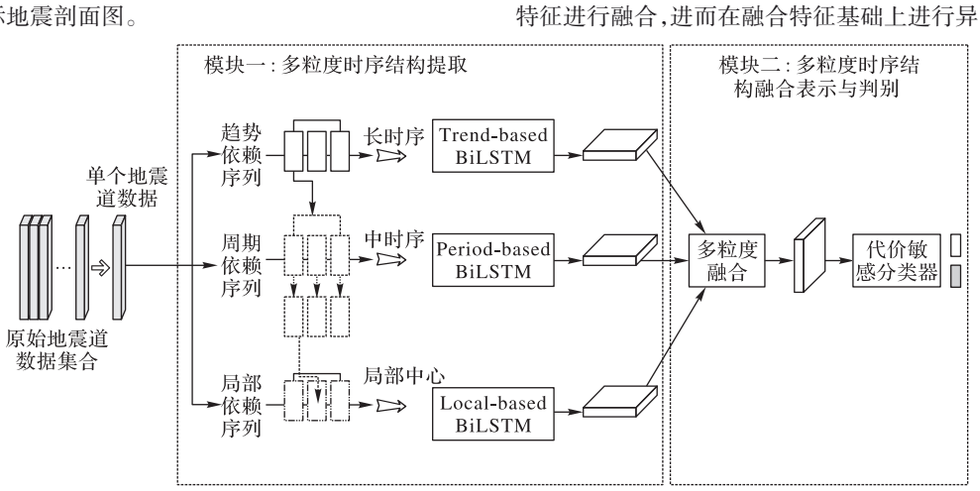

图 2　本文算法总体流程

Fig. 2　Overall flow of the proposed algorithm

www.joca.cn

3.1 多粒度时序结构特征抽取模块

原始地震道数据是通过地震波进行采集的，其内部的时序性可从三个不同的时序粒度进行刻画。三个时序粒度分别是指地震道数据内长时序呈现出的趋势性、中时序呈现的周期性以及局部关键点周围具有的局部性特征。通过以上三个粒度可从全局至中观、再到局部以不同层次对原始地震资料中的潜在特征进行学习和表示，进而表达出地震道数据内部存在的时序结构。具体而言过程如下。

1）趋势性结构特征。针对地下岩层存在复杂结构的特点，单个地震道数据 $\boldsymbol{x}_i$ 隐式地包含了地质岩层渐变的规律和特点。例如不同岩性的岩层从地质结构上往往呈现出一定潜在渐变过程。因此首先将单个地震道数据和岩性标记切分多个长时间序列 $\boldsymbol{x}_i=\{\boldsymbol{x}_{it_1},\boldsymbol{x}_{it_2},\cdots,\boldsymbol{x}_{it_n}\}$ 作为训练样本，以捕捉给定区域呈现的趋势性特点。将其通过趋势性结构特征模型（Trend-LSTM，T-LSTM）进行训练，可得到该粒度下的特征表示 $\boldsymbol{h}_{it_k}\in\mathbf{R}^n$ 以及 T-LSTM 模型参数。给定单个地震道样本 $(\boldsymbol{x}_i,\boldsymbol{y}_i)$，特征提取过程如式（2）：

$$\boldsymbol{h}_{it_k}=f_{\mathrm{T}}(\boldsymbol{x}_{it_k},\boldsymbol{W}_{\mathrm{t}},\boldsymbol{b}_{\mathrm{t}}) \tag{2}$$

其中：$f_{\mathrm{T}}(\cdot)$ 表示 T-LSTM 模型计算函数，$\boldsymbol{x}_{it_k}$ 表示第 i 个地震道数据中第 t_k 个长序列，$\boldsymbol{W}_{\mathrm{t}}$ 和 $\boldsymbol{b}_{\mathrm{t}}$ 表示 T-LSTM 的参数。

T-LSTM 采用经典 BiLSTM 作为编码器产生隐含层的编码表示。后续两个粒度的时序结构提取模型与该处一致。

2）周期性结构特征。由于地震波在传输过程中受波长和特定地层结构等物理因素影响，所以本文在得到单个地震道数据长序列片段 $\boldsymbol{x}_{it_k}$ 基础上，进一步结合所给定区域地质特性，将已得到的趋势性长序列进一步划分为中序列，用以捕获中观特性。即 $\boldsymbol{x}_{it_k}=\left\{x_{it_k}^{(1)}\cup x_{it_k}^{(2)},\cdots,\cup x_{it_k}^{(m)}\right\}$，其中 m 为该长序列片段中依据波长周期等特性进行划分后的中时序个数。通过将其输入至周期性结构特征模型（Period-LSTM，P-LSTM）进行训练，将第 q 个中时序的地震道子序列形成独立特征表示。通过式（3）可获得第 q 个中时序的特征表示 $\boldsymbol{h}_{\mathrm{ip}}^{(q)}\in\mathbf{R}^n$。其中，$f_{\mathrm{P}}(\cdot)$ 表示 P-LSTM 模型计算函数，$\boldsymbol{W}_{\mathrm{p}}$ 和 $\boldsymbol{b}_{\mathrm{p}}$ 表示 P-LSTM 模型参数。在得到中时序每个特定波长表示 $\boldsymbol{h}_{ip}^{(q)}$ 基础上将其进行拼接，即可得到给定长序列的周期性结构特征 $\boldsymbol{h}_{ip}^{(q)}\in\mathbf{R}^n$。

$$\boldsymbol{h}_{ip_k}^{(q)}=f_{\mathrm{P}}(\boldsymbol{x}_{it_k}^{(q)},\boldsymbol{W}_{\mathrm{p}},\boldsymbol{b}_{\mathrm{p}}) \tag{3}$$

3）局部性结构特征。为进一步捕捉地质结构中岩层局部存在的差异性，本文在考虑趋势性（长时序）和周期性（中时序）基础上，对单个地震道特定波长内关键局部点序列进行开窗提取，用以获得局部性特征。即通过局部性结构特征模型（Local-LSTM，L-LSTM）对局部结构进行学习，得到局部关键点时序结构的独立特征表示 $\boldsymbol{h}_{il}\in\mathbf{R}^n$。具体而言，首先从给定中时序中选择中心点进行局部子序列开窗提取，构成 $\boldsymbol{L}_{it_k}^{q}$ 集合。其次，将该集合通过 L-LSTM 模型进行训练，通过式（4）提取含有局部结构的隐含特征表示：

$$\boldsymbol{h}_{il}^{(q)}=f_{\mathrm{L}}(\boldsymbol{L}_{it_k}^{(q)},\boldsymbol{W}_{\mathrm{l}},\boldsymbol{b}_{\mathrm{l}}) \tag{4}$$

其中：$f_{\mathrm{L}}(\cdot)$ 表示 L-LSTM 的计算函数，$\boldsymbol{W}_{\mathrm{l}}$ 和 $\boldsymbol{b}_{\mathrm{l}}$ 表示 L-LSTM 模型参数。在得到每一个局部表示 $\boldsymbol{h}_{il}^{(q)}$ 基础上，将其进行拼接，即可得到给定长序列的局部性结构特征 $\boldsymbol{h}_{il}\in\mathbf{R}^n$。

3.2 多粒度时序结构特征融合与判别模块

在获得多个粒度独立特征后，本文进一步对不同粒度的特征进行自动化融合，即对给定地震道长时序包含的趋势性、周期性和局部性三种不同粒度的结构特征进行融合，形成多粒度时序结构特征 $\boldsymbol{H}_i$，本文采用的融合方法如式（5）。其中，$\widehat{\boldsymbol{W}}_{\mathrm{t}}$、$\widehat{\boldsymbol{W}}_{\mathrm{p}}$ 和 $\widehat{\boldsymbol{W}}_{\mathrm{l}}$ 分别表示了趋势性结构特征 $\boldsymbol{h}_{it}$、周期性结构特征 $\boldsymbol{h}_{ip}$ 和局部性结构特征 $\boldsymbol{h}_{il}$ 的融合参数。在训练阶段，本文采用端到端的方式进行联合训练，即模型中的参数 $\theta_\Omega=(\widehat{\boldsymbol{W}}_{\mathrm{t}},\widehat{\boldsymbol{W}}_{\mathrm{p}},\widehat{\boldsymbol{W}}_{\mathrm{l}})$ 在初始化后，通过误差回传的方式进行端到端参数学习。

$$H_i=\Psi(\boldsymbol{h}_{it},\boldsymbol{h}_{ip},\boldsymbol{h}_{il}|\theta_\Omega)=\widehat{\boldsymbol{W}}_{\mathrm{t}}\circ\boldsymbol{h}_{it}+\widehat{\boldsymbol{W}}_{\mathrm{p}}\circ\boldsymbol{h}_{ip}+\widehat{\boldsymbol{W}}_{\mathrm{l}}\circ\boldsymbol{h}_{il} \tag{5}$$

为了进一步利用已勘测井段监督信息对多粒度特征存在的代价敏感问题进行处理，本文设计了对融合后的特征采用代价敏感的全连接分类器进行学习，其要求分类后的样本与原样本含油性标记保持一致性的基础上，考虑到假阳性样本和假阴性样本代价敏感影响。式（6）表示将融合后的多粒度特征通过全连接分类器进行含油性判别：

$$\hat{Y}_i=F(\boldsymbol{H}_i|\theta_\Omega)=f_D(\boldsymbol{H}_i,\boldsymbol{W}_D,\boldsymbol{b}_D) \tag{6}$$

3.3 算法模型参数优化和损失函数

本文算法模型参数和损失分别可从两个阶段进行分析：第一个阶段是提取时序结构时模型涉及的参数和损失；第二阶段是融合各独立特征和对该特性进行含油性检测判别时模型参数和损失。在本文算法训练过程，采用了随机梯度下降的方法进行参数优化，具体由以下两个部分组成：

1）多粒度时序结构融合参数和损失。在多粒度时序结构提取阶段，模型参数涉及包括 3 个部分：分别是趋势性结构模型参数、周期性结构模型参数和局部性结构模型参数。其中，$f_{\mathrm{T}}(\cdot)$、$f_{\mathrm{P}}(\cdot)$ 和 $f_{\mathrm{L}}(\cdot)$ 分别表示 3 个粒度特征提取模型。$(\boldsymbol{W}_{\mathrm{t}},\boldsymbol{b}_{\mathrm{t}},\boldsymbol{W}_{\mathrm{p}},\boldsymbol{b}_{\mathrm{p}},\boldsymbol{W}_{\mathrm{l}},\boldsymbol{b}_{\mathrm{l}})$ 分别表示该阶段待优化的模型参数。

2）基于多粒度融合表示代价敏感参数和损失。获得多粒度融合特征后，将其输入基于代价敏感的全连接分类器进行判别模型，$\widehat{\boldsymbol{W}}$ 表示该阶段待优化的模型参数，$f_D(\cdot)$ 表示含油性预测模型。为了有效地区分有油和无油样本，本文算法将损失函数建模为有监督异常检测模型。因此，损失函数在设计时加大了对假阳性样本和假阴性样本错误计算造成的损失，该基于代价敏感型损失函数如式（7）：

$$\begin{aligned}L(\hat{Y}_i,Y_i,\boldsymbol{C})=&Y_i*(c_1*\log(\hat{Y}_i)+c_2*\log(1-\hat{Y}_i)+\\&(1-Y_i)*(c_3*\log(1-\hat{Y}_i)+c_4*\log(\hat{Y}_i))\end{aligned} \tag{7}$$

其中 $\boldsymbol{C}=(c_1,c_2,c_3,c_4)$ 为代价敏感向量，本文代价敏感向量由物探领域专家结合江苏油田实际地质特性进行研判并作为模型先验领域知识。

4 实验与结果分析

本章将本文算法在江苏油田实际地震资料上进行了验证。

4.1 数据集描述

为进一步验证本文所提出算法在真实地震资料上的预测结果，本次实验在江苏油田实际地震资料数据集上进行实验。由于江苏地区油田分布的自然局限性，导致钻井数据本身稀少。本次实验数据集以真永联三维地震资料为基础，该区域位于苏北盆地高邮凹陷深凹带。目前该区已发现真武、许庄、曹庄、肖刘庄和永安等多个油气田，成藏条件较为有利，是江苏油田主要产油区。该工区共拥有地震道数据 221 121 条，经

www.joca.cn

过实际测井验证,共有81条含测井标记的地震道数据,其中含油性地震道数据为20条。经预处理后,实验得到该工区地震道数据中含标记信息480个时序点,其中含油标记点共有31个(单个地震道可存在多个含油性标记点),有监督的信息总量占比较少且类别不平衡。

4.2　对比方法

在实验方法对比方面,本文算法将分别与以下两类算法进行对比:

1)与油气检测领域所采用的机器学习算法对比。主要包括:神经网络(BackPropagation, BP)算法[20]、支持向量机(Support Vector Machine, SVM)算法[22]、卷积神经网络(Convolutional Neural Network, CNN)算法[28]和LSTM算法[30]。

2)与机器学习领域经典浅层和深度算法进行比对实验,其中浅层算法包括k最近邻(k-Nearest Neighbor, kNN),深度算法包括门控循环单元(Gated Recurrent Unit, GRU)网络和循环神经网络(Recurrent Neural Network, RNN)算法。

从以上两个方面进行对比,方法选择一方面因为是物探领域研究团队在探索机器学习技术时较常用的经典方法;另一方面,也是进一步在真实地震资料上验证和探索本文所提出算法与其他经典机器学习方法在性能上的差异性。此外,关于比对算法的超参数设置,本文在实验时对江苏油田永真联地震资料进行了验证集划分,通过其在验证集上的性能比较,选择了最佳超参数进行设置并开展后续实验与评估。

4.3　实验设置

本文所采用的单道地震波数据共包含6 000个振幅值,形式如图3所示。其中左侧为靠近地面的地震波数据,右侧是远离地面的地震波数据。经分析可发现,左侧振幅值的疏密程度高于右侧,主要原因是地震波自地面往地下传导的过程中存在着衰减。根据对江苏油田真永联区域总共包含81条含油标记地震波数据的分析,其中含油数据20条,无油数据61条。同时,油层分布较薄,最小出现的含油标记点位置为2 041,最大的含油标记点位置为2 545,除2 545点外含油点位于1 500~2 500。为有效地学习出其中特征,结合物探领域专家建议,本次实验的趋势序列、周期序列和局部序列之间的特征分别按如下规则取出数据:

1)趋势性序列划分:重点针对1 000~4 000的振幅值进行划分。

2)周期性序列划分:重点针对1 500~3 000的振幅值进行划分。

3)局部性序列划分:重点针对2 000~2 500的振幅值进行划分。

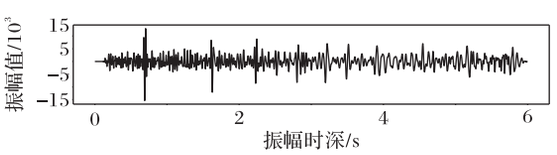

图3　单道地震道可视化振幅

Fig. 3　Visualized amplitude of single seismic trace

本次实验评价主要采用以下四个评价指标:

1)召回率(Recall)表示预测正确的含油性地震道数据与实际地震资料中所有含油性地震道数据的比率,依据真阳性(True Positive, TP)、假阳性(False Positive, FP)、真阴性(True Negative, TN)、假阴性(False Negative, FN)进行计算,过程如式(8)所示:

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

2)准确率(Accuracy)表示预测正确的含油性地震道数据与预测出所有含油性地震道数据的比率,计算过程如(9):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

3) ROC(Receiver Operating Characteristic)曲线下方的面积(Area Under Curve, AUC)大小表示对预测算法的曲线下面积, AUC值越大的分类器,正确率越高。

4)收敛性表示判断所提出算法和其他对比方法在永真联三维地震资料中进行训练是否可经过多步迭代之后趋于某一固定值。

4.4　实验结果与分析

本次实验结果从四个方面对本文算法进行评价。其中本文算法与对比算法(含加上不平衡代价设计)的召回率、准确率和AUC分别如表2和表3所示,收敛性分析如图4所示。

表2　不同算法的召回率、精度和AUC结果对比

Tab. 2　Comparison of different algorithms on Recall, Precision and AUC

分类	算法	召回率	准确率	AUC
基于浅层的算法	kNN	0.00	0.65	0.42
	BP	0.00	0.71	0.59
	SVM	0.25	**0.76**	0.46
基于深度的算法	RNN	0.83	0.25	0.63
	LSTM	0.83	0.25	0.63
	GRU	0.65	0.00	0.42
	CNN	0.00	0.76	0.50
	本文算法(去代价敏感)	0.50	**0.82**	0.71
	本文算法	**1.00**	0.59	**0.73**

表3　不同算法代价敏感性能对比结果

Tab. 3　Comparison results of cost-sensitive performance of different algorithms

分类	召回率	准确率	AUC
RNN+代价敏感	0.25	0.52	0.43
LSTM+代价敏感	**1.00**	0.23	0.50
GRU+代价敏感	**1.00**	0.23	0.50
本文算法	**1.00**	**0.59**	**0.73**

1)召回率分析。表2结果表明:在经典非时序类算法模型大多过拟合于无油,无法识别出有油的样本,导致召回率较低。其中kNN、BP与CNN将测试集中所有的有油数据预测为无油,SVM识别出少量的含油数据。基于时序的算法RNN, GRU和LSTM在性能上明显高于其他经典算法,表明捕捉数据内部时序的重要性。本文提出的算法能够捕捉多粒度的特征,在召回率上取得了最佳性能,且优于对比算法的平均性能(平均性能召回率为0.36)。表3结果表明:将不平衡代价损失引入到经典深度模型中,召回率相较之前有所提升。

2)准确率分析。表2结果表明:本文方法准确率优于所对比算法的平均性能(平均性能准确率为0.48),但未取得最佳性能。经分析,由于无油数据在数据集中占据大多数的比例(类别不平衡),除本文方法之外其他对比模型大多将测试样本直接预测为无油,所以使得本文方法的预测准确率低于

www.joca.cn

部分经典方法。表3结果表明:将不平衡代价损失引入到经典深度模型中,准确率相较之前稍有下降。

3)AUC分析。表2结果表明:本文算法在AUC指标上获得了最佳性能,相对于第二名RNN和LSTM算法,得到了10%提高。此外,本文提出的算法优于平均性能(平均性能AUC为0.52)。表3结果表明:将不平衡代价损失引入到经典深度模型中,虽然召回率有所提高,但AUC结果并未发现有所提升。因此,未发现引入代价敏感策略对于原有经典深度学习方法性能有本质提升。

4)收敛性分析。本文提出算法和所对比的4类深度学习算法收敛性结果如图4所示,从收敛性结果上而言,本文算法和对比方法能够在给定地震资料上逐渐降低训练损失。同时,根据收敛性结果进一步对比了本文方法与LSTM和GRU算法的训练耗时,本文方法耗时约是其三倍。由于本文算法采用3个BiLSTM作为特征提取器,耗时与单个算法模型训练耗时基本保持线性增加关系。

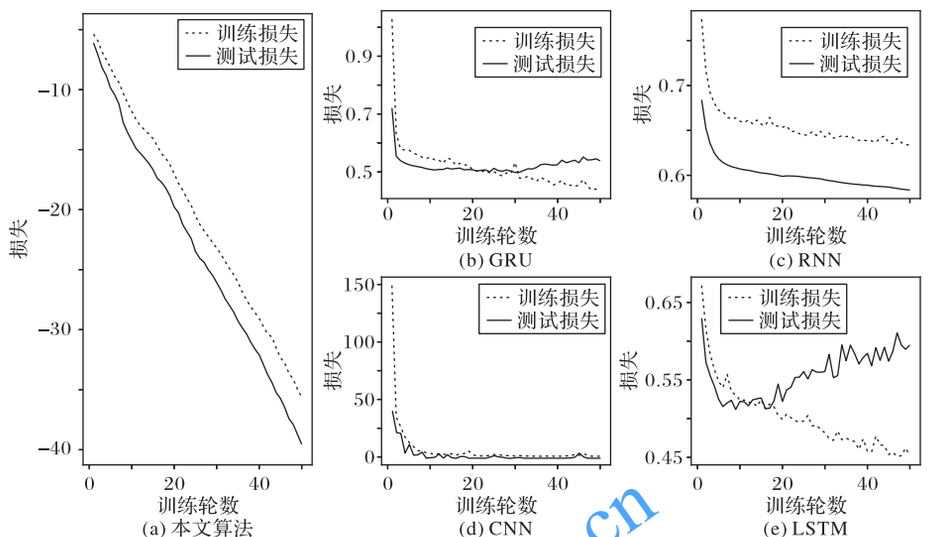

图4 深度学习算法收敛性对比

Fig. 4 Convergence comparison of different deep learning based algorithms

5 结语

对地下储层含油性勘测是石油勘探领域热点研究方向之一。本文针对地下储层含油性检测问题进行了探索,提出了一种基于多粒度时序结构表示的异常检测算法。该方法首先对于单个地震道数据提取多粒度时序结构并形成独立特征;其次,在提取多个粒度时序结构基础上进行特征融合,形成对地震道数据的综合表示;最后,通过对融合后特征进行联合训练与判别,得到对于该地震数据含油性检测结果。本文算法在苏北盆地高邮凹陷深凹带实际地震资料上进行了实验仿真,实验结果表明:本文算法相对比传统机器学习方法能够更有效地进行储层含油性预测。

地下岩层随着深度和广度的不断增加,其地质结构也随之存在潜在变化,如何更加有效地捕捉地震道数据内和数据之间存在的潜在分布变化或概念漂移现象,并将其特征融入含油性判别任务中是值得进一步探索的研究内容。

参考文献(References)

[1] GHOLAMI R, MORADZADEH A, RASOULI V, et al. Shear wave velocity prediction using seismic attributes and well log data[J]. Acta Geophysica, 2014, 62(4):818-848.

[2] GRAVES A, JAITLY N, MOHAMED A R. Hybrid speech recognition with deep bidirectional LSTM[C]// Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. Piscataway: IEEE, 2013: 273-278.

[3] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7):1527-1554.

[4] SARIKAYA R, HINTON G E, DEORAS A. Application of deep belief networks for natural language understanding[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(4):778-784.

[5] LeCUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436-444.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6):84-90.

[7] KIROS J R, CHAN W, HINTON G E. Illustrative language understanding: large-scale visual grounding with image search[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2018:922-933.

[8] MENG F D, CHEN S Y, ZHANG Y C, et al. Characterization of motor oil by laser-induced fluorescence [J]. Analytical Letters, 2015, 48(13): 2090-2095.

[9] CUDJOE S, VINASSA M, HENRIQUE BESSA GOMES J, et al. A comprehensive approach to sweet-spot mapping for hydraulic fracturing and CO_2 huff-n-puff injection in Chattanooga shale formation [J]. Journal of Natural Gas Science and Engineering, 2016, 33:1201-1218.

[10] EBRAHIMI A, KHAMEHCHI E. Developing a novel workflow for natural gas lift optimization using advanced support vector machine [J]. Journal of Natural Gas Science and Engineering, 2016, 28: 626-638.

[11] ZHANG J G. Application of gray Elman neural network to predict the gas emission amount [J]. Advanced Materials Research, 2013, 706/707/708:1750-1754.

[12] 罗浩然,尹成,丁峰,等. 概率神经网络的平滑参数分析及在地

www.joca.cn

震属性分析中的应用[J]. 石油物探, 2017, 56(4): 551-558. (LUO H R, YIN C, DING F, et al. The smoothing parameter analysis of probabilistic neural network and its application in seismic attribute analysis [J]. Geophysical Prospecting for Petroleum, 2017, 56(4): 551-558.)

[13] MASOUDI P, TOKHMECHI B, JAFARI M A, et al. Application of fuzzy classifier fusion in determining productive zones in oil wells [J]. Energy Exploration and Exploitation, 2012, 30(3): 403-415.

[14] 王振洲,张春雷,高世臣. 利用决策树方法识别复杂碳酸盐岩岩性——以苏里格气田苏东41-33区块为例[J]. 油气地质与采收率, 2017, 24(6):25-33. (WANG Z Z, ZHANG C L, GAO S C. Lithology identification of complex carbonate rocks based on decision tree method: an example from Block Sudong41-33 in Sulige Gas Field [J]. Petroleum Geology and Recovery Efficiency, 2017, 24(6):25-33.)

[15] LI Q, ZHONG H Q, WANG Y, et al. Integrated development optimization model and its solving method of multiple gas fields [J]. Petroleum Exploration and Development, 2016, 43(2): 293-300.

[16] 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 X Y, YE D N, ZHANG G Z. 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.)

[17] 刘瑞林,马在田. 神经网络在油气评价和预测方面的研究现状[J]. 地球物理学进展, 1995, 10(2): 75-84. (LIU R L, MA Z T. The current studies of application of neural networks to oil and gas evaluation and prediction[J]. Progress in Geophysics, 1995, 10(2): 75-84.)

[18] 陆文凯,牟永光. 利用BP神经网络进行测井资料外推[J]. 石油地球物理勘探, 1996, 31(5): 712-715. (LU W K, MOU Y G. Logging data extrapolation using BP neural network [J]. Oil Geophysical Prospecting, 1996, 31(5): 712-715.)

[19] 张奏,郑晓东,李劲松,等. 基于SOM和PSO的非监督地震相分析技术[J]. 地球物理学报, 2015, 58(9): 3412-3423. (ZHANG Y, ZHENG X D, LI J S, et al. Unsupervised seismic facies analysis technology based on SOM and PSO[J]. Chinese Journal of Geophysics, 2015, 58(9): 3412-3423.)

[20] 张向君,李幼铭,刘洪. 神经网络结构风险最小油气预测[J]. 石油地球物理勘探, 2002, 37(1):73-76. (ZHANG X J, LI Y M, LIU H. Oil and gas prediction by neural networks with structure-risk-minimum[J]. Oil Geophysical Prospecting, 2002, 37(1):73-76.)

[21] MOHEBBI A, KAMALPOUR R, KEYVANLOO K, et al. The prediction of permeability from well logging data based on reservoir zoning, using artificial neural networks in one of an Iranian heterogeneous oil reservoir [J]. Petroleum Science and Technology, 2012, 30(19): 1998-2007.

[22] 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[J]. 石油地球物理勘探, 2012, 47(2): 282-285. (ZHANG C K, JIANG X D, ZHU Z Y, et al. Attributes selection and reservoir prediction based on support vector machine[J]. Oil Geophysical Prospecting, 2012, 47(2):282-285.)

[23] 王波,夏同星,谭辉煌. 基于斑块饱和模型井控属性融合法油气检测[J]. 石油物探, 2017, 56(2):288-294. (WANG B, XIA T X, TAN H H. The well-controlled attributes fusion method for hydrocarbon detection based on patchy-saturation model [J]. Geophysical Prospecting for Petroleum, 2017, 56(2):288-294.)

[24] 袁照威,陈龙,高世臣,等. 基于马尔可夫-贝叶斯模拟算法的多地震属性沉积相建模方法——以苏里格气田苏区块为例[J]. 油气地质与采收率, 2017, 24(3):37-43. (YUAN Z W, CHEN L, GAO S C, et al. A method of sedimentary facies modeling through integration of multi-seismic attributes based on Markov-Bayes model: an example from Su10 area in the north of Sulige gas field[J]. Petroleum Geology and Recovery Efficiency, 2017, 24(3):37-43.)

[25] 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[J]. 石油地球物理勘探, 2016, 51(6):1202-1211. (SONG J G, GAO Q S, LI Z. Application of random forests for regression to seismic reservoir prediction [J]. Oil Geophysical Prospecting, 2016, 51(6):1202-1211.)

[26] SAIKIA P, BARUAH R D, SINGH S K, et al. Artificial neural networks in the domain of reservoir characterization: a review from shallow to deep models[J]. Computers and Geosciences, 2020, 135: No. 104357.

[27] 林年添,付超,张栋,等. 无监督与监督学习下的含油气储层预测[J]. 石油物探, 2018, 57(4):601-610. (LIN N T, FU C, ZHANG D, et al. Supervised learning and unsupervised learning for hydrocarbon prediction using multiwave seismic data [J]. Geophysical Prospecting for Petroleum, 2018, 57(4):601-610.)

[28] 林年添,张栋,张凯,等. 地震油气储层的小样本卷积神经网络学习与预测[J]. 地球物理学报, 2018, 61(10): 4110-4125. (LIN N T, ZHANG D, ZHANG K, et al. Predicting distribution of hydrocarbon reservoirs with seismic data based on learning of the small-sample convolution neural network[J]. Chinese Journal of Geophysics, 2018, 61(10): 4110-4125.)

[29] 付超,林年添,张栋,等. 多波地震深度学习的油气储层分布预测案例[J]. 地球物理学报, 2018, 61(1): 293-303. (FU C, LIN N T, ZHANG D, et al. Prediction of reservoirs using multi-component seismic data and the deep learning method [J]. Chinese Journal of Geophysics, 2018, 61(1): 293-303.)

[30] 安鹏,曹丹平,赵宝银,等. 基于LSTM循环神经网络的储层物性参数预测方法研究[J]. 地球物理学进展, 2019, 34(5): 1849-1858. (AN P, CAO D P, ZHAO B Y, et al. Reservoir physical parameters prediction based on LSTM recurrent neural network[J]. Progress in Geophysics, 2019, 34(5): 1849-1858.)

MENG Fan, born in 1988, Ph. D. candidate, senior engineer. His research interests include machine learning, unsupervised outlier detection.

CHEN Guang, born in 1992, M. S. candidate. His research interests include data mining.

WANG Yong, born in 1962, Ph. D., professor of engineering. His research interests include seismic data processing.

GAO Yang, born in 1970, Ph. D., professor. His research interests include big data analysis, machine learning, multi-agent system, video/image processing.

GAO Dequn, born in 1970, M. S., senior engineer. His research interests include petroleum geological exploration.

JIA Wenlong, born in 1984, engineer. His research interests include big data, artificial intelligence.

www.joca.cn