

王倩楠, 王治国, 杨阳等. 2024. 基于多特征融合自编码器的无监督地震相分类研究. 地球物理学报, 67(1):370-378, doi:10.6038/cjg2023Q0871.

Wang Q N, Wang Z G, Yang Y, et al. 2024. Unsupervised seismic facies classification based on multi-feature fusion autoencoder. *Chinese J. Geophys.* (in Chinese), 67(1):370-378, doi:10.6038/cjg2023Q0871.

基于多特征融合自编码器的无监督地震相分类研究

王倩楠^{1,2}, 王治国^{1,2*}, 杨阳^{2,3}, 朱剑兵⁴, 高静怀^{2,3}

1 西安交通大学数学与统计学院, 西安 710049

2 陕西国家应用数学中心, 西安 710049

3 西安交通大学信息与通信工程学院, 西安 710049

4 中国石油化工股份有限公司胜利油田分公司物探研究院, 东营 257022

摘要 地震相分类是地震数据解释中的一个重要步骤, 是地震数据与沉积相的连接工具. 为了提高地震相分类精度和减少对有限人工标签的依赖, 本文提出了一种基于多特征融合自编码器的无监督地震相分类方法. 首先, 提出了一种混合卷积和变分编码的多特征融合自编码器, 实现了地震数据中表征地震相的大量隐含特征提取. 其次基于非负矩阵分解和 K 均值聚类实现了主特征分量分解和地震相聚类. 实际地震数据应用结果和指标分析表明, 本文方法提取的隐含特征趋于正态分布, 且主特征分量中蕴含了不同地震相类别的响应, 从而可以获得更准确的地震相分类结果. 在渤海湾盆地东营凹陷古近系沙河街组湖相沉积中, 清晰划分出了六类沉积微相的边界, 有利于揭示三角洲沉积环境演变.

关键词 地震相分类; 多特征融合自编码器; 卷积自编码器; 变分自编码器; 非负矩阵分解

doi:10.6038/cjg2023Q0871

中图分类号 P631

收稿日期 2022-10-31, 2023-02-07 收修定稿

Unsupervised seismic facies classification based on multi-feature fusion autoencoder

WANG QianNan^{1,2}, WANG ZhiGuo^{1,2*}, YANG Yang^{2,3}, ZHU JianBing⁴, GAO JingHui^{2,3}

1 School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China

2 National Center for Applied Mathematics in Shaanxi, Xi'an 710049, China

3 School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China

4 Geophysical Research Institute, Shengli Oilfield Company, Sinopec, Dongying 257022, China

Abstract Seismic facies classification plays an important role in seismic data interpretation, which is a bridge between seismic data and sedimentary facies. To improve the accuracy of seismic facies classification and reduce the number of manual labels, we propose an unsupervised seismic facies classification method based on the Multi-Feature Fusion Autoencoder (MFAE). At first, the proposed MFAE is generated by the hybride convolution network and variational autoencoder, which extract a large number of latent variables from seismic data. Then, the nonnegative matrix factorization is utilized to implement the principal eigencomponent decomposition and the K -means clustering is also introduced to obtain the results of seismic clustering. Finally, the proposed method is applied to the real data to test the performance. The results reveal that the extracted features of the proposed methods approximately satisfy the Gaussian distribution and the main

基金项目 国家自然科学基金(41974137)资助.

第一作者简介 王倩楠, 女, 2000 年生, 硕士研究生, 研究方向为地震智能解释. E-mail: QN.Wang@stu.xjtu.edu.cn

* 通讯作者 王治国, 男, 1982 年生, 教授, 研究方向为勘探地球物理. E-mail: emailwzg@mail.xjtu.edu.cn

features after principal eigencomponent decomposition contains the responses of different seismic facies classes. Therefore, the accuracy of the seismic facies classification can be improved. For the Paleogene Shahejie formation in Dongying Sag, the proposed method can predict more clear boundaries of the six seismic facies classes, which is beneficial to demonstrating the evolution of the deltaic sedimentary.

Keywords Seismic facies classification; Multi-Feature Fusion Autoencoder (MFAE); Convolutional autoencoder; Variational autoencoder; Non-negative matrix factorization

0 引言

地震相是特定沉积相的地震响应。地震相分类,是指人工或自动的提取和分类不同的地震响应特征(即地震属性),实现不同地震相的划分,进而解释出对应的地下沉积环境和岩相分布等(Dumay and Fournier, 1988; Nivlet, 2007)。传统的人工地震解释中,地震相分类高度依赖于解释人员的丰富经验和大量时间投入,但其解释结果会带有一定的主观性(Saggaf et al., 2003; Wrona et al., 2018)。因此,伴随机器学习的不断发展,基于机器学习自动实现地震相分类,一直是地震解释中一个重要的研究方向(穆星, 2005; Chandrashekar and Sahin, 2014; 解滔等, 2016; Zhao et al., 2018; Liu et al., 2020; Wang et al., 2020; Puzyrev and Elders, 2022; 许志远等, 2022; 硕良勋等, 2023; 翟亮等, 2023)。

根据是否依赖于已知地震相类别标签,可以将基于机器学习的地震相分类分为有监督学习和无监督学习两大类。监督学习以优选的地震属性作为输入,并以标注的地震相作为训练标签,完成地震相的划分。例如,Alaudah 等(2018)利用相似性检索与弱监督学习实现地震相标签制作,减少人工制作标签数据的耗时,但划分的地震相标签精度不高。Kim 等(2019)使用非线性统计和随机森林决策树算法分析了多个地震属性与地震相之间的相关性,构建出了一个有效属性子集来实现地震相分类。Qi 等(2020)通过高斯混合模型搜索反映地质响应的最佳地震属性集来实现地震相分类。然而,在有监督学习中,已知地震相类别标签需要有经验解释人员划分,导致了地震相标签的数量总是受限的;同时区分某一目标相的最佳地震属性也在一定程度上依赖于有经验解释人员的选择。因此,为避免分类结果受人工标注地震相标签的影响,降低地震属性的主观性选择,使用无监督学习实现地震相分类是另一条重要途径(Barnes, 2007; Marroquín, 2014; 张奕等,

2015; Zhao et al., 2018; Qi et al., 2020)。例如,de Matos 等(2007)使用小波变换识别每个地质信息段中的地震道奇点,然后使用自组织聚类得到地震相。张奕等(2015)基于自组织映射和粒子群优化的 K 均值聚类实现优选属性的地震相自动识别。总之,无论是有监督学习还是无监督学习,如何更好的提取地震响应特征,是提高地震相分类精度的关键和难点。

近年来,深度学习已经成为了机器学习中最主流的方向。卷积自编码器(Convolutional AutoEncoder, CAE)和变分自编码器(Variational AutoEncoder, VAE)都是基于深度学习的特征提取算法,已经被成功用于地震数据特征提取和地震相分类。如 Qian 等(2018)、Puzyrev 和 Elders(2022)利用卷积自动编码器对地震数据进行特征表示学习,然后利用 K 均值算法实现地震相聚类。Li 等(2020)将带标签的数据放入变分自编码器提取隐含特征,并使用高斯混合模型拟合每类标签数据的特征分布,将拟合的特征分布放入解码器实现地震相预测。但是,两类自编码器在地震相分类中也有一些不足:卷积自编码器提取的隐含特征具有输入数据的空间特征,但不具有统计分布规律,不利于地震相定性解释;变分自编码器能够提取服从目标分布的隐含特征,但忽略了特征空间的表示,使得提取的特征不能直接与地震相空间位置关系建立联系。

为融合以上两类自编码器各自的优点,本文提出了一种多特征融合自编码器(Multi-Feature Fusion AutoEncoder, MFAE)用于提取地震数据的隐含特征,使提取的隐含特征能够统计地震相类别,更有利于地震相定性解释。MFAE 具有一种专门设计的编码器-解码器结构。其创新之处在于:编码器能够同时进行卷积和变分两种编码;解码器既能接收卷积自编码器的隐含特征,也能接收变分自编码器的隐含特征;MFAE 设计了联合损失函数,使重构输出数据不失真且提取的隐含特征趋于标准正态分布。在 MFAE 提取多个地震响应特征后,为了减

少数据储存空间和缩短聚类时间,采用非负矩阵分解(Non-negative Matrix Factorization, NMF)和 K 均值聚类进行了大量隐含特征的主特征提取和聚类,从而获得了准确的地震相分类,并成功用于东营凹陷三维地震数据的地震相解释.

1 原理及方法

1.1 多特征融合自编码器

1.1.1 网络结构

MFAE 由两部分组成(图 1),第一部分是编码器,对输入地震数据做降采样,获得大量隐含特征.第二部分是解码器,对获得的大量隐含特征做上采样,输出最后的重构地震数据.将地震数据 x 作为输入送入 MFAE 的编码器,其编码器 f_k 主要由两个部分组成,上部分为卷积自编码器 f_{kcae} ,下部分为变分自编码器 f_{kvae} .将经过卷积自编码器 f_{kcae} 提取的隐含特征 Z_{CAE} 与经过变分自编码器 f_{kvae} 提取的隐含特征 Z_{VAE} 加权求和,得到最终的隐含特征 $Z_{CAE+VAE}$.MFAE 的解码器 g_k 输入也由两个部分组成,第一个部分为 Z_{CAE} ,第二个部分为 $Z_{CAE+VAE}$.将 Z_{CAE} 和 $Z_{CAE+VAE}$ 送入解码器的第一个解码层 g_{k-1} 分别得到两个解码结果 $g_{k-1}(Z_{CAE})$ 与 $g_{k-1}(Z_{CAE+VAE})$,将两个解码结果加权求和后,一起送入解码器剩下的解码层 g_k 得到重构的地震数据 \hat{x} ,表达式为

$$f_k = f_{kcae} + f_{kvae}, \tag{1}$$

$$Z_{CAE} = f_{kcae}(x), \tag{2}$$

$$Z_{VAE} = f_{kvae}(x), \tag{3}$$

$$Z_{CAE+VAE} = \mu Z_{CAE} + (1 - \mu) Z_{VAE}, \tag{4}$$

$$\hat{x} = g_k(\lambda g_{k-1}(Z_{CAE}) + (1 - \lambda) g_{k-1}(Z_{CAE+VAE})). \tag{5}$$

1.1.2 损失函数

损失函数值由 MFAE 的解码输出地震数据与编码输入地震数据共同决定.通过最小化损失函数,使预测的输出地震数据接近于输入地震数据,使编码器能正确地对地震数据进行编码、隐含特征能正确表示输入地震数据、解码器能够利用隐含特征进行完整的重构.因 MFAE 融合了 CAE 和 VAE 的网络结构特征,其联合损失函数 L_{AE} 主要由 CAE 的重构损失函数——均方误差 L_{MSE} 和 VAE 的 KL 散度损失函数 L_{KL} 组成. MFAE 通过最小化均方误差损失函数 L_{MSE} 使 MFAE 能够更好地重构输入地震数据,使输出地震数据 \hat{x} 和输入地震数据 x 保持较高的相似度.在 VAE 部分,通过均值和方差计算 KL 散度损失函数 L_{KL} ,使大量隐含特征的分布 P 能够服从预设的分布 Q .MFAE 为了让输出地震数据 \hat{x}_i 更完整地重构输入地震数据 x_i ,MFAE 定义了新的重构损失函数 L_{rec} , L_{rec} 在均方误差损失函数 L_{MSE} 的基础上加入了 SmoothL1 损失 $L_{SmoothL1}$,使得 MFAE 在进行地震数据重构时,增加重构地震图像内部不同值边界两侧的光滑性,表达式为

$$L_{MSE} = \frac{\sum_{i=1}^n (\hat{x}_i - x_i)^2}{n}, \tag{6}$$

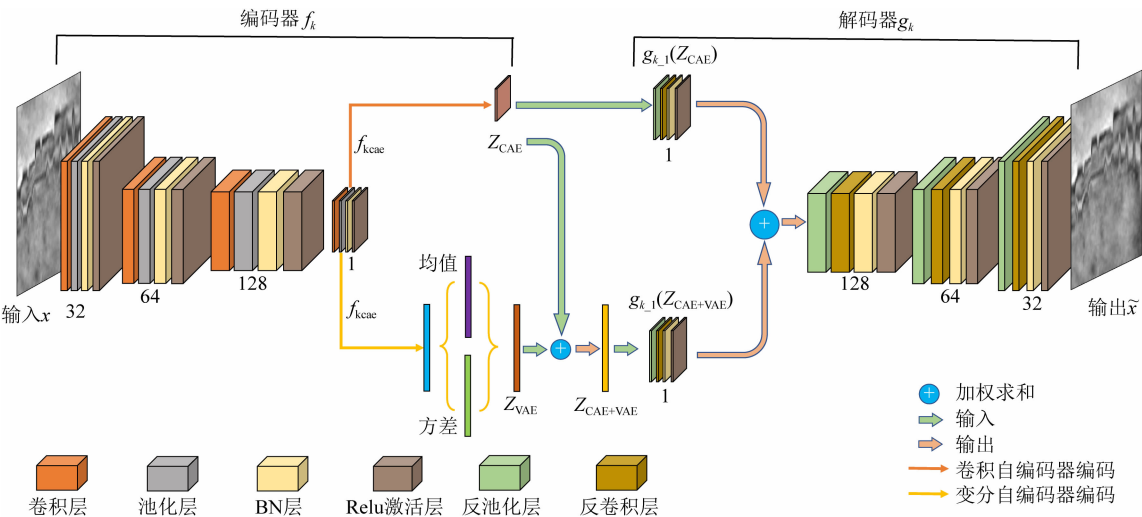


图 1 多特征融合自编码器网络结构

在编码器中,卷积层的输出通道数分别为 32、64、128、1;在解码器中,反卷积层的输入通道数分别为 1、128、64、32.

Fig. 1 The structure of MFAE

In the encoder, the number of output channels of the convolution layers are 32, 64, 128, 1, respectively. In the decoder, the number of input channels of the deconvolution layers are 1, 128, 64, 32, respectively.

$$L_{\text{SmoothL1}} = \begin{cases} 0.5(\tilde{x}_i - x_i)^2, & \text{if } |\tilde{x}_i - x_i| < 1, \\ |\tilde{x}_i - x_i| - 0.5, & \text{otherwise.} \end{cases} \quad (7)$$

$$L_{\text{rec}} = \eta L_{\text{SmoothL1}} + L_{\text{MSE}}, \quad (8)$$

式中 \tilde{x}_i 是预测值, x_i 是真实值, $\eta \in (0, 1)$ 是一个平衡 L_{MSE} 和 L_{SmoothL1} 的超参数, 用以控制惩罚力度。

假设两个分布 P, Q 的概率密度函数为 $p(x_i)$ 和 $q(x_i)$, 定义两个分部之间的 KL 散度损失函数 L_{KL} 如公式(9)所示:

$$L_{\text{KL}} = \text{KL}(P \parallel Q) = \sum_i p(x_i) \lg \frac{p(x_i)}{q(x_i)}, \quad (9)$$

联合损失函数 L_{AE} 为

$$L_{\text{AE}} = (1 - \alpha)L_{\text{rec}} + \alpha L_{\text{KL}}, \quad (10)$$

式中 α 是权重系数, $\alpha \in (0, 1)$ 取不同的值可以得到不同的模型效果。一般来说, 重构损失函数占的权重越大, 解码器输出的结果越接近真实的地震数据, 但

提取的大量隐含特征的分布就越远离标准正态分布; KL 散度占的权重越大, 提取的大量隐含特征的分布越接近标准正态分布, 但是解码器输出的地震数据越可能失真。为了解决上述问题, 本文使解码器接收卷积自编码器的隐含特征的同时也能接收变分自编码器的隐含特征, 如图 1 和公式(5)。公式(5)平衡了重构损失函数 L_{rec} 和 KL 散度 L_{KL} 二者所占权重的问题, 既解决了输出地震数据失真问题, 又使得大量隐含特征分布趋于标准正态分布。

1.2 无监督地震相分类

利用 MFAE 获得大量隐含特征后, 利用非负矩阵分解减少冗余的隐含特征并提取前三个主特征分量, 再利用 K 均值聚类方法对地震相进行分类, 方法流程如图 2 所示。

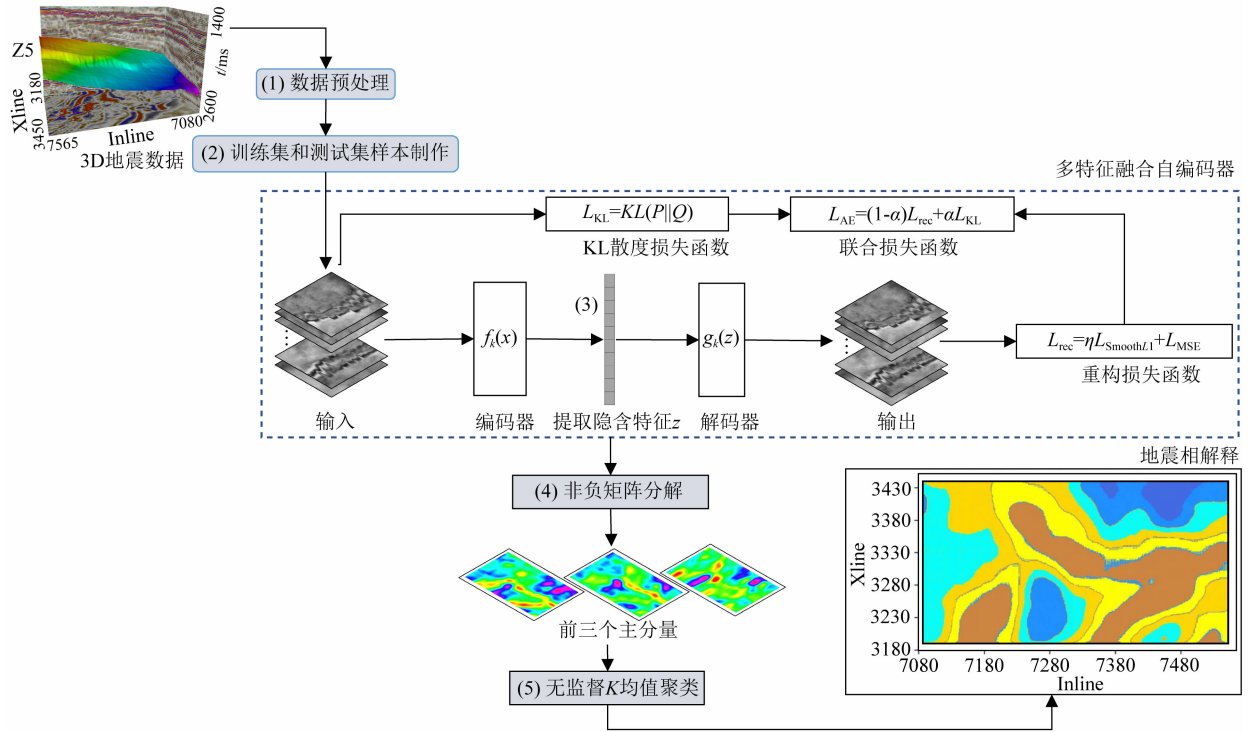


图 2 无监督地震相分类的流程

Fig. 2 Workflow of the proposed unsupervised seismic facies classification

(1)地震数据预处理;在进行自编码器训练前,应对地震数据进行预处理,将地震数据的值转化为 0 到 1 之间,使得地震数据可以视为地震剖面灰度地震图像。从而地震相的分类就对应了灰度图的像素值分类。这样做的目的是能够降低网络训练的难度,提高重构地震数据的准确度。预处理公式为

$$X_{\text{nor}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (11)$$

其中 $X_{\text{nor}} \in R^{a \times b}$ 是预处理后的值, $X \in R^{a \times b}$ 是原始地震数据, X_{\min} 和 X_{\max} 分别是地震数据中的最小值和

最大值, a 和 b 分别表示地震图像的大小。

(2)制作训练编码器的训练集和测试集,送入编码器进行训练和测试;具体制作方法如图 3 所示,采用滑动时窗,依次在地震图像上以每个像素点为中心,以 $\frac{p}{2}$ 的滑动步长截取大小为 $p \times p$ 的图像块 x_i , 并用补零的方式填补地震图像边缘的图像块空值。然后将得到的地震数据图像块随机划分为训练集和测试集。

(3)将地震数据输入训练好的自编码器,提取大量隐含特征;假设大小为 $x_i \in R^{p \times p}$ 的地震图像块经

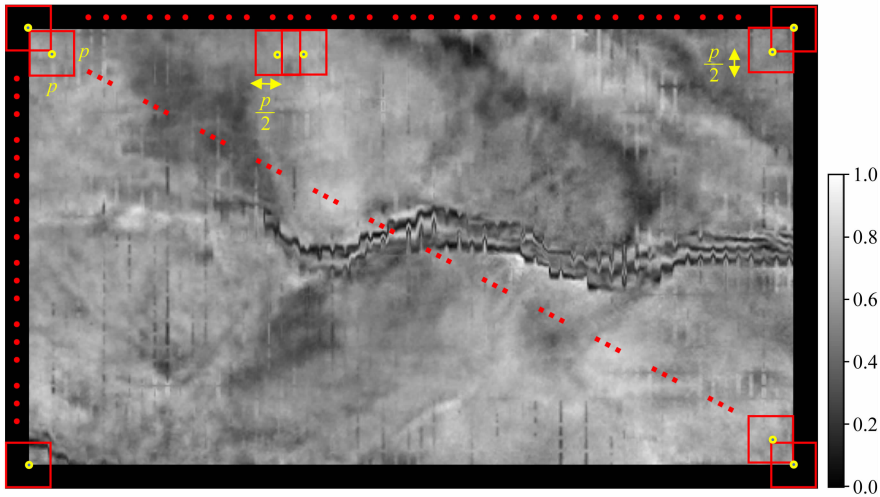


图 3 基于地层切片的样本数据集制示意图

Fig. 3 Diagram of the sample dataset based on stratigraphic slices

过编码器处理后变为了大小为 $x'_i \in R^{1 \times p'}$ 的向量. 那么将整个地震剖面数据 X_{nor} 经过编码器 f_k 处理后, 得到了大量隐含特征, 并组成集合 Z :

$$Z = f_k(X_{\text{nor}}), \quad (12)$$

$X_{\text{nor}} \in R^{a \times b}$ 是预处理后的值, $Z \in R^{n \times p'}$ (其中, n 的大小为 $n = a \times b$) 是获得的隐含特征集合, a 和 b 分别表示地震图像的大小.

(4) 利用非负矩阵(NMF), 将大量冗余特征压缩为小样本特征, 得到所有隐含特征的前三个主特征分量; 在对所有的隐含特征 Z 进行非负矩阵分解前, 需利用公式(13)进行归一化处理, 将所有隐含特征变为非负隐含特征 Z' , 再进行分解处理:

$$Z'_{n \times p'} \approx W_{n \times 3} \cdot V_{3 \times m}. \quad (13)$$

(5) 对提取的所有隐含特征的前三个主特征分量 $W_{n \times 3}$ 进行无监督 K 均值聚类, 最后得到地震相分类图 \tilde{X} :

$$\tilde{W}_{n \times 1} = \text{Kmeans}(W_{n \times 3}), \quad (14)$$

$$\tilde{X}_{a \times b} = \text{Reshape}(n \times 1, a \times b)(\tilde{W}_{n \times 1}), \quad (15)$$

其中 $\text{Kmeans}(\cdot)$ 表示 K 均值聚类算法, 公式(15)表示将输入数据 $\tilde{W}_{n \times 1}$ 的大小由 $R^{n \times 1}$ 变为 $R^{a \times b}$ 并获得最终的地震相图 $\tilde{X}_{a \times b}$.

基于 MFAE 的无监督地震相分类代码如算法 1 所示.

2 实际数据应用

2.1 研究区概况

研究区位于东营凹陷内部, 主力储层处于古近系沙河街组沉积时期. 沿东营凹陷长轴方向是三角

算法 1 基于 MFAE 的无监督地震相分类

输入: 沿层位的 RMS 振幅属性值 $X \in R^{a \times b}$, 最大迭代次数 n .

过程:

1: 基于 $X_{\text{nor}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$ 计算归一化后的 RMS 振幅属性值 $X_{\text{nor}} \in R^{a \times b}$, 并划分训练集 X_{n1} 和测试集 X_{n2} , $X_n = \{x_1, x_2, \dots, x_i, \dots, x_n\}$, $x_i \in R^{p \times p}$;

2: 将 X_n 送入自编码器 MFAE 进行训练与测试

for $i = 1:n$

计算编码结果 $Z_i = f_k(x_i)$, 解码结果 $\hat{x}_i = g_k(Z_i)$, $Z_i \in R^{1 \times p'}$, $\hat{x}_i \in R^{p \times p}$

根据式(6)、(7)、(8)、(9)、(10)计算损失 L_{AE}

if L_{AE} 收敛 then

break

end

保存训练好的编码结构 f'_k ;

3: 将 $X_{\text{nor}} \in R^{a \times b}$ 送入 f'_k , 提取所有隐含特征 $Z = f'_k(X_{\text{nor}})$, $Z \in R^{n \times p'}$;

4: 将所有隐含特征变为非负隐含特征 $Z' = \frac{Z - Z_{\min}}{Z_{\max} - Z_{\min}}$, $Z' \in R^{n \times p'}$, 并利用非负矩阵分解提取前三个主特征分量 $W_{n \times 3}$, $Z'_{n \times p'} \approx W_{n \times 3} \cdot V_{3 \times m}$;

5: 对 $W_{n \times 3}$ 进行 K 均值聚类 $\tilde{W}_{n \times 1} = \text{Kmeans}(W_{n \times 3})$, 得到地震相聚类结果 $\tilde{X}_{a \times b}$, $\tilde{X}_{a \times b} = \text{Reshape}(n \times 1, a \times b)(\tilde{W}_{n \times 1})$;

输出: 沿层位的地震相聚类结果 $\tilde{X}_{a \times b}$.

洲发育的鼎盛时期, 随着三角洲不断地自东向西进积, 研究区在沙三段晚期已全部被三角洲所覆盖. 研究区在三角洲前缘半深湖—深湖地带发育了许多浊积岩体, 岩性以长石细砂岩、岩屑细砂岩及粉砂岩为主, 浊流沉积是油气藏形成的有利条件, 是油气勘探的重要对象(Zhang, 2004). 研究区目的层为图 4 中的 Z5 层位, Z5 层位由下至上经历了三角洲前缘

沉积过渡至三角洲平原沉积. 该时期沉积于湖盆发育早中期, 构造运动强烈, 盆地周围一些大型河流三角洲开始出现, 导致湖水进一步淡化.

2.2 地震相分类结果

针对 Z5 层位的地震属性输入数据, 设置公式 (4)、(5)、(8) 里的 μ 为 0.5, λ 为 0.5, η 为 0.125 时, 能够使 MFAE 较好地重构地震输入图像且使损失函数收敛. 沿 Z5 层位提取沿层位切片的均方根振幅属性值 $X \in R^{271 \times 485}$, 并对沿层切片的属性值进行归一化处理. 将归一化处理后的沿层切片送入

MFAE 得到所有隐含特征 $Z \in R^{131435 \times 676}$, 并对所有的隐含特征进行非负矩阵分解, 提取蕴含有效信息最多的前三个主特征分量 $Z' \in R^{131435 \times 3}$, 如图 5 所示. 将提取的所有隐含特征的前三个主特征分量送入 K 均值聚类得到最终的地震相分类结果, 如图 6 所示.

结合分析 Z5 层位聚类结果与该层位所处沉积时期, 可得出该层位处形成了西北、西南、西三个方向的三条水下分支流河道, 三角洲前缘在明显扩大, 河道的两侧形成分支流河口沙坝, 东部物源体系继

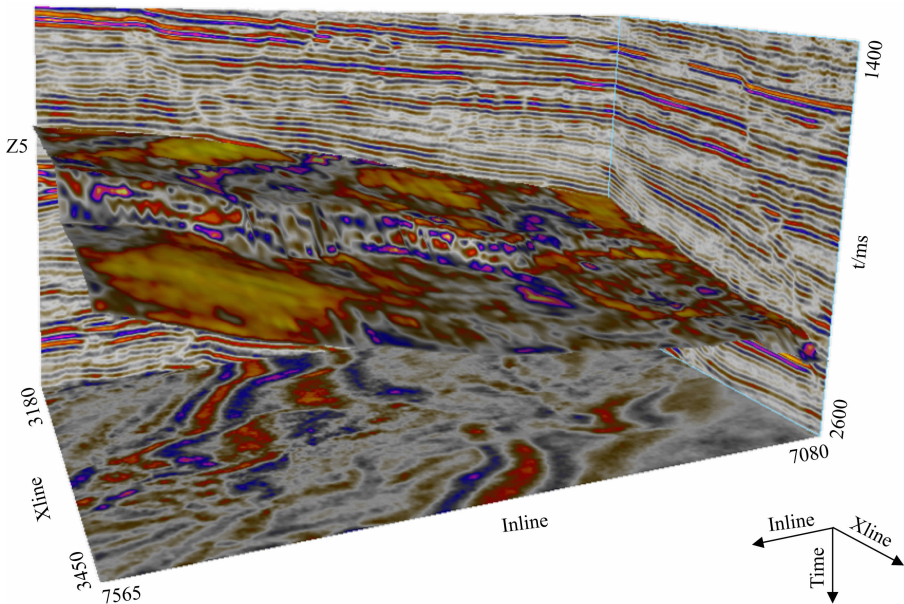


图 4 研究区的三维地震数据及其 Z5 层位
Fig. 4 3D seismic data and the horizon Z5 in this study area

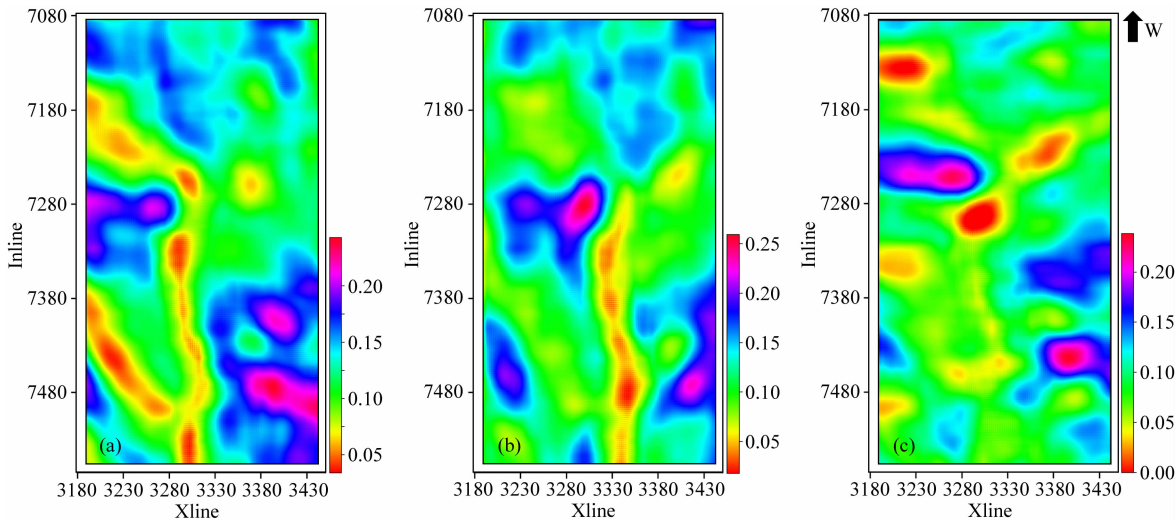


图 5 所有隐含特征的前三个主特征分量
(a) 第一个分量; (b) 第二个分量; (c) 第三个分量.

Fig. 5 The first three components of the latent variables

(a) The first principal component; (b) The second principal component; (c) The third principal component.

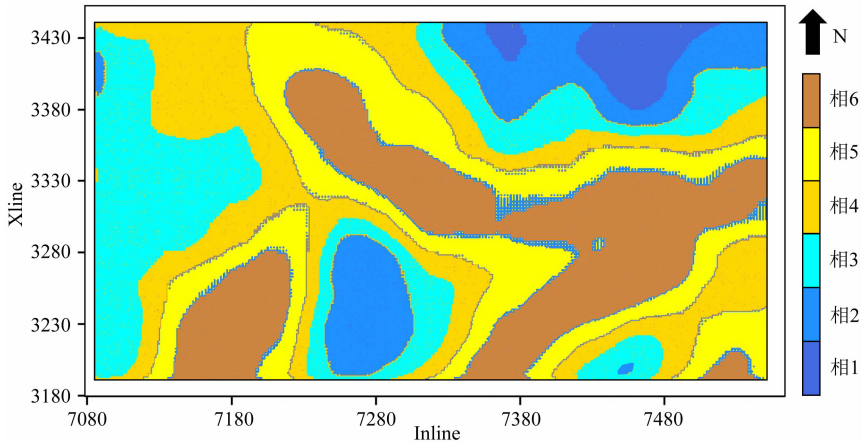


图 6 沿 Z5 层位的无监督地震相分类结果

Fig. 6 Clustering results of seismic attributes at Z5 horizon

续向西推进. 根据湖泊-三角洲的沉积模式, 可解释出相 1 为半深湖、相 2 为浅湖、相 3 为滨湖、相 4 为远砂坝、相 5 为分支流河口沙坝、相 6 为水下分支流河道. 由于河道发生淤塞, 河道两侧积水后形成湖泊, 因此, 分支流河口沙坝(相 5)与为水下分支流河道(相 6)之间形成了少量的浅湖相(相 2). 在 Z5 层位聚类结果的右下角, 因为水流对河道的冲刷与侵蚀, 致使河流愈来愈弯曲, 最后导致河流自然截弯取直, 使得原来弯曲的河道废弃, 形成牛轭湖. 因湖水淡化, 牛轭湖最终变为滨湖(相 3)与浅湖(相 2). 经过上述对 Z5 层位的沉积相分析, 得出三角洲不断地自东向西进积, 湖水在进一步淡化的结论, 符合研究区的沉积演化规律.

3 讨论与分析

为了进一步分析 MFAE 提取地震数据特征上的优势, 下面讨论不同自编码器所提取隐含特征的概率密度分布和聚类结果的量化指标.

首先, 综合对比 CAE、VAE、MFAE 三种自编码提取的所有隐含特征频率分布直方图(图 7)可知, 利用 CAE 提取的所有隐含特征不符合正态分布, VAE 提取的所有隐含特征完全符合正态分布, MFAE 提取的所有隐含特征能够消除 CAE 提取的异常分布隐含特征, 且使提取的所有隐含特征趋于正态分布.

在没有真实地震相分类标签的情况下, 采用轮廓系数(Rousseeuw, 1987)评估聚类效果. 轮廓系数通过计算一个簇的紧致度, 用以评价 K 均值聚类对聚类结果所产生的影响. 其值介于 $[-1, 1]$, 值越大代表同类别样本距离越相近, 不同类别样本距离越

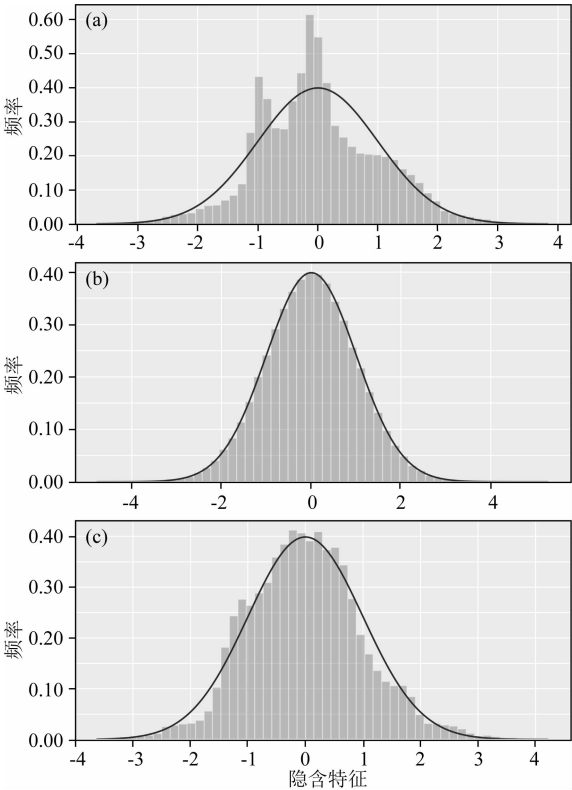


图 7 沿 Z5 层位(a)CAE、(b)VAE 和(c)MFAE 提取的隐含特征的频率分布直方图

黑线表示拟合的正态分布.

Fig. 7 Frequency distribution histogram of Z5 horizon latent features extracted by (a) CAE, (b) VAE, and (c) MFAE

The black lines represent the fitted normal distribution.

远, 内聚度和分离度都相对较优. 为了说明 MFAE 能够提取有效的隐含特征, 实验将六种不同方式提取的所有隐含特征, 用于 K 均值聚类, 并对比评估它们的聚类效果, 如表 1 所示.

表 1 中的 VAE 表示将 VAE 提取的所有隐含特征直接用于 K 均值聚类; VAE + NMF 表示将

表 1 基于不同隐含特征的聚类结果评估

Table 1 Evaluation of clustering results based on different latent features

聚类的 输入数据	VAE	CAE	MFAE	VAE+ NMF	CAE+ NMF	MFAE+ NMF
轮廓系数	0.03	0.10	0.12	0.32	0.44	0.56

VAE 提取的所有隐含特征先进行非负矩阵分解,再进行 K 均值聚类;其他同组实验同理。对比分析 MFAE、CAE、VAE 三组实验的轮廓系数,发现在 CAE 与 VAE 的特征融合后,MFAE 更有利于后续进行聚类处理。对比后三组实验,发现经过非负矩阵分解处理后的隐含特征,其聚类效果均优于直接用编码器提取的隐含特征。其中,MFAE 与 NMF 组合后,其聚类结果的轮廓系数达到了六组实验中的最高值(0.56),说明本文提出的 MFAE 方法在地震相分析中是有效的,且该方法有力提高了地震相分类的准确性。

4 结论

(1)本文提出了一种基于多特征融合的自编码器 MFAE。MFAE 是一种特殊的编码-解码结构,融合了卷积编码和变分编码实现地震数据的隐含特征提取。输出的大量隐含特征趋于正态分布,且不同特征中蕴含了清晰的不同地震相类别的响应。

(2)本文实现了一种基于 MFAE 的无监督地震相分类方法。首先 MFAE 能够提取地震数据中潜在的大量隐含特征;其次非负矩阵分解能够减少冗余,得到突出不同类别地震相的主特征分量;最终 K 均值聚类获得地震相分类。实际应用效果和指标分析表明,本文提出方法可以提供一个更准确的无监督地震相分类结果。

(3)在渤海湾盆地东营凹陷的地震相分类应用中,清晰划分出了古近系沙河街组半深湖、浅湖、滨湖、远砂坝、分支流河口沙坝、水下分支流河道等六类沉积微相,揭示了三角洲不断地自东向西进积,湖水在进一步淡化的沉积环境演变。

致谢 感谢中国石化胜利油田公司提供本文使用的地震数据。

References

Alaudah Y, Alfarraj M, Alregib G. 2018. Structure label prediction using similarity-based retrieval and weakly supervised label

mapping. *Geophysics*, 84(1): V67-V79, doi: 10.1190/geo2018-0028.1.

Barnes A E. 2007. Redundant and useless seismic attributes. *Geophysics*, 72(3): P33-P38, doi: 10.1190/1.2716717.

Chandrashekar G, Sahin F. 2014. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1): 16-28, doi: 10.1016/j.compeleceng.2013.11.024.

deMatos M C, Osorio P L, Johann P R. 2007. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. *Geophysics*, 72(1): P9-P21, doi: 10.1190/1.2392789.

Dumay J, Fournier F. 1988. Multivariate statistical analyses applied to seismic facies recognition. *Geophysics*, 53(9): 1151-1159, doi: 10.1190/1.1442554.

Kim Y, Hardisty R, Marfurt K J. 2019. Attribute selection in seismic facies classification: application to a Gulf of Mexico 3D seismic survey and the Barnett Shale. *Interpretation*, 7(3): SE281-SE297, doi: 10.1190/int-2018-0246.1.

Li K H, Chen S, Hu G M. 2020. Seismic labeled data expansion using variational autoencoders. *Artificial Intelligence in Geosciences*, 1: 24-30, doi: 10.1016/j.aig.2020.12.002.

Liu M L, Jervis M, Li W C, et al. 2020. Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial networks. *Geophysics*, 85(4): O47-O58, doi: 10.1190/geo2019-0627.1.

Marroquin I D. 2014. A knowledge-integration framework for interpreting seismic facies. *Interpretation*, 2(1): SA1-SA9, doi: 10.1190/int-2013-0057.1.

Mu X. 2005. Method of seismic facies auto-classification by the seismic geometrical attribute and the self-organizing neural network. *Geological Science and Technology Information (in Chinese)*, 24(3): 109-112, doi: 10.3969/j.issn.1000-7849.2005.03.022.

Nivlet P. 2007. Uncertainties in seismic facies analysis for reservoir characterisation or monitoring: causes and consequences. *Oil & Gas Science and Technology-Revue d'IFPEN*, 62(2): 225-235, doi: 10.2516/ogst:2007019.

Puzyrev V, Elders C. 2022. Unsupervised seismic facies classification using deep convolutional autoencoder. *Geophysics*, 87(4): IM125-IM132, doi: 10.1190/geo2021-0016.1.

Qi J, Zhang B, Lyu B, et al. 2020. Seismic attribute selection for machine-learning-based facies analysis. *Geophysics*, 85(2): O17-O35, doi: 10.1190/geo2019-0223.1.

Qian F, Yin M, Liu X Y, et al. 2018. Unsupervised seismic facies analysis via deep convolutional autoencoders. *Geophysics*, 83(3): A39-A43, doi: 10.1190/geo2017-0524.1.

Rousseeuw P J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20: 53-65, doi: 10.1016/0377-0427(87)90125-7.

Saggaf M M, Toksöz M N, Marhoon M I. 2003. Seismic facies classification and identification by competitive neural networks. *Geophysics*, 68(6): 1984-1999, doi: 10.1190/1.1635052.

Shuo L X, Zhao Y H, Chai B F, et al. 2023. Semi-supervised adversarial

network for seismic facies classification. *Progress in Geophysics* (in Chinese), 38(5): 2105-2113, doi: 10.6038/pg2023GG0372.

Wang Z G, Gao D L, Lei X L, et al. 2020. Machine learning-based seismic spectral attribute analysis to delineate a tight-sand reservoir in the Sulige gas field of central Ordos Basin, Western China. *Marine and Petroleum Geology*, 113: 104136, doi: 10.1016/j.marpetgeo.2019.104136.

Wrona T, Pan I, Gawthorpe R L, et al. 2018. Seismic facies analysis using machine learning. *Geophysics*, 83(5): O83-O95, doi: 10.1190/geo2017-0595.1.

Xie T, Zheng X D, Zhang Y. 2016. Seismic facies analysis based on linear prediction cepstrum coefficients. *Chinese Journal of Geophysics* (in Chinese), 59(11): 4266-4277, doi: 10.6038/cjg20161127.

Xu Z Y, Xiong X J, Xiao Y, et al. 2022. Reservoir prediction of Sinian Deng 4 member in the intra-platform of Moxi area based on seismic texture attributes clustering method. *Progress in Geophysics* (in Chinese), 37(3): 1207-1213, doi: 10.6038/pg2022FF0248.

Zhai L, Zhang Q, Chen Y R, et al. 2023. Facies constrained inversion method and application based on lithology compensated coefficient; a case study of the southern part of Chengdao Oilfield in Bohai Bay Basin. *Progress in Geophysics* (in Chinese), 38(5): 2350-2359, doi: 10.6038/pg2023GG0566.

Zhang S W. 2004. The application of an integrated approach in exploration of lacustrine turbidites in Jiyang Sub-Basin, Bohai Bay Basin, China. *Journal of Petroleum Science and Engineering*, 41(1-3): 67-77, doi: 10.1016/s0920-4105(03)00144-x.

Zhang Y, Zheng X D, Li J S, et al. 2015. Unsupervised seismic

facies analysis technology based on SOM and PSO. *Chinese Journal of Geophysics* (in Chinese), 58(9): 3412-3423, doi: 10.6038/cjg20150933.

Zhao T, Li F Y, Marfurt K J. 2018. Seismic attribute selection for unsupervised seismic facies analysis using user-guided data-adaptive weights. *Geophysics*, 83(2): O31-O44, doi: 10.1190/geo2017-0192.1.

附中文参考文献

穆星. 2005. 利用地震几何属性和自组织神经网络进行地震相的自动识别. *地质科技情报*, 24(3): 109-112, doi: 10.3969/j.issn.1000-7849.2005.03.022.

硕良勋, 赵云鹤, 柴变芳等. 2023. 基于半监督对抗网络的地震相识别. *地球物理学进展*, 38(5): 2105-2113, doi: 10.6038/pg2023GG0372.

解滔, 郑晓东, 张葵. 2016. 基于线性预测倒谱系数的地震相分析. *地球物理学报*, 59(11): 4266-4277, doi: 10.6038/cjg20161127.

许志远, 熊晓军, 肖尧等. 2022. 基于地震纹理属性聚类分析的磨溪地区台内带灯四段储层预测研究. *地球物理学进展*, 37(3): 1207-1213, doi: 10.6038/pg2022FF0248.

翟亮, 张强, 陈永芮等. 2023. 一种基于岩性补偿系数的沉积相约束反演方法及应用研究——以渤海湾盆地埕岛南部东营组为例. *地球物理学进展*, 38(5): 2350-2359, doi: 10.6038/pg2023GG0566.

张葵, 郑晓东, 李劲松等. 2015. 基于 SOM 和 PSO 的非监督地震相分析技术. *地球物理学报*, 58(9): 3412-3423, doi: 10.6038/cjg20150933.

(本文编辑 张正峰)