

# 基于响应曲线聚类的时间域电磁法 地电结构类型划分研究

施明智<sup>1,2</sup>, 曹辉<sup>1</sup>, 陈柏午<sup>1</sup>

1. 成都理工大学 地球物理学院, 成都 610059;  
2. 南宁职业技术学院 人工智能学院, 南宁 530008

[摘要] 为了充分利用时间域电磁法正演响应曲线携带的特征信息实现地电结构类型快速预测, 通过将曲线的分布特性和形态特性相结合构造了曲线距离, 采用相关系数作为权重参数衡量曲线形态相似程度, 实现了时间域电磁法正演响应曲线聚类。应用效果表明, 依据该方法得到的聚类结果能实现正演响应样本的标签化, 从而通过多分类任务神经网络完成地电结构的类型划分。

[关键词] 时间域电磁法; 曲线聚类; 地电结构

[分类号] P319.2

[文献标志码] A

## Classification of geoelectric structures using time-domain electromagnetic method based on response curve clustering

SHI Mingzhi<sup>1,2</sup>, CAO Hui<sup>1</sup>, CHEN Baiwu<sup>1</sup>

1. College of Geophysics, Chengdu University of Technology, Chengdu 610059, China;

2. School of Artificial Intelligence, Nanning College for Vocational Technology, Nanning 530008, China

**Abstract:** In this study, we aimed to make full use of the characteristic information carried by the forward response curve in the time-domain electromagnetic method to quickly and accurately predict the types of geoelectric structures. We developed the curve distance by combining the characteristics of distribution and morphology of the curve, and used the correlation coefficient as the weighting parameter to measure its degree of similarity. This enabled us to cluster the forward response curve of the time-domain electromagnetic method. The results of its application showed that the outcomes of clustering obtained by this method could be used to label the forward response samples, and to subsequently classify geoelectric structures by using a multi-classification neural network.

[收稿日期] 2023-04-24。

[基金项目] 国家自然科学基金(41974090)。

[第一作者] 施明智(1983—), 男, 高级工程师, 博士, 研究方向: 时间域电磁法数值模拟中的机器学习, E-mail: 170132357@qq.com。

[通信作者] 曹辉(1974—), 男, 教授, 博士, 研究方向: 电磁法物理模拟与数值模拟, E-mail: Caohui@cdut.edu.cn。

[引用格式] 施明智, 曹辉, 陈柏午, 2024. 基于响应曲线聚类的时间域电磁法地电结构类型划分研究[J]. 成都理工大学学报(自然科学版), 51(3): 522-530.

Shi M Z, Cao H, Chen B W, 2024. Classification of geoelectric structures using time-domain electromagnetic method based on response curve clustering[J]. Journal of Chengdu University of Technology (Science & Technology Edition), 51(3): 522-530.

**Key words:** time-domain electromagnetic method; curve clustering; geoelectric structure

时间域电磁法观测的是随时间变化的电磁场传播特性,其正演响应对应不同的地下电性结构会体现出不同的变化特征。这种突出的特征对应关系能够极大发挥出深度学习技术对于特征学习和特征提取的优势。聚类是一种典型的无监督深度学习技术,其擅长数据挖掘任务,能通过提取数据的特征来分析出数据的潜在意义,可用于地球物理数据解释。聚类有助于对数据特征进行归纳,展现数据间内在结构的相关性,由此来完成对数据的自动归类(Song et al., 2010; Qiang et al., 2010)。在地球物理领域,聚类常用来实现电磁数据识别、降噪或信噪分离(赵玄等, 2018; 张贤等, 2022),以及地质构造的识别和归类(黄颖等, 2019)等。

通过聚类的方法对时间域电磁法正演响应曲线进行特征提取,就能实现对地电结构模型的类型划分,而划分结果可以作为类型标签用于构建多分类任务神经网络的训练样本数据集。为了能够有效地将曲线的分布特性和形态特性相结合,充分利用曲线携带的特征信息,得到更为准确的聚类结果,首先需要定义符合时间域电磁法正演响应曲线特征的距离,然后采用该距离进行曲线相似性度量,并构造能用于衡量曲线间形态相似程度的系数,从而采用加权的方法实现对时间域电磁法正演响应曲线的聚类,得到机器对地电结构模型样本集的类型划分结果。

## 1 响应曲线聚类

聚类算法常用于对数据进行分簇和自动归类,旨在按照一定的分簇条件,从全部数据中找到符合条件的数据进行归簇。各簇数据作为密切相关的观测结果具有高度相似的特征,且同簇数据比他簇数据更加相似。常见的聚类方法多用于对二维平面或三维空间上的点进行分组,这些算法显然不能直接移植到对平面中曲线的分簇上。与对点的聚类不同,对曲线进行聚类需要解决的问题有:(1)对样本数据进行向量形式的构造;(2)对曲线之间距离定义的构造;(3)对曲线相似性度量及判定准则的构造。

电磁勘探得到的正演响应曲线反映的是地下电性结构的分布对电磁场的感应变化。对正演响

应曲线进行聚类的目的,是为了探究一簇正演响应曲线与一种类型的地下电性结构的对应关系。为了找到适合正演曲线聚类的方法,需要结合正演曲线的特点,从聚类距离的计算方法、聚类中心选取、聚类迭代过程等方面进行对比和选择。

曲线聚类虽不同于对点的聚类,但由于其分簇划定的量化标准主要还是依靠对距离的度量,因此可将其归于划分聚类。基于划分的聚类通常是以将样本距离作为相似性度量为特征的一类聚类方法,其中采用欧几里得距离(Euclidean distance, 又称为欧氏距离)进行距离度量的 K-Means 算法(Hartigan and Wong, 1976),是一种应用广泛的划分聚类方法。由于其采用欧几里得距离作为相似性判定准则,因此不但适合二维平面上的点,也能适应像正演曲线这样的欧几里得空间中的高维向量。K-Means 聚类采用 Kaufman 法(Kaufman and Rousseeuw, 1990)确定聚类初始中心点,通过不断寻找最合适的聚类中心来实现最小化平方误差函数的求取,从而得到  $K$  个聚集了最理想的划分结果的簇,划分后的簇中每一个样本距离该簇中心都近于其他簇中心。

作为无监督深度学习技术的一种,聚类对误差函数最优解空间的搜索往往具有随机性,高维度空间里选择合适的距离度量标准相当困难(胡静, 2007)。对曲线进行聚类时,如果仅考虑欧几里得距离的远近作为曲线相似性度量的唯一标准,会由于较高维度的样本特征而容易产生等距离的数据样本(吴亚雄等, 2020),且由于忽略曲线的轮廓特征和形态变化特征等外部特征,影响了聚类效果的进一步提高,存在一定的局限性(朱晓清, 2012; 吴亚雄等, 2020; 党倩等, 2022)。K-Means 聚类算法仅体现了对距离的相似性判别,不能完全反映曲线形态和变化趋势,需要加以改进才能适用于正演曲线聚类。

### 1.1 曲线向量化

聚类的一个关键问题是对样本的距离进行合理度量,聚类算法中对距离的不同定义会导致不同的聚类结果。K-Means 算法通过计算样本之间的距离来完成对样本相似度的量化,从而为样本分簇提供依据。相对于平面上的点,二维平面上曲线距离的度量远比点的度量要复杂,依据不

同的目的可采用多样的距离计算方法。

聚类问题中对距离的度量,与地球物理资料反演问题中采用不同的范数来完成对长度的度量有相通之处。地球物理反演问题可以理解为从数据空间到模型空间的映射问题,如果解存在,则模型构建的过程即是通过模型参数进行多次修正迭代来求解一个与观测数据达成最佳拟合的模型的过程。在模型构建中,采用范数来对测量长度进行定义,以此来求解测度最小的模型参数。在各种反演方法中,对测度的不同定义导致了模型参数和观测数据的加权不相同。也就是说,由于不同范数产生了不同加权,从而产生了构建模型的差异化,并导致观测数据或模型参数的统计量体现出不同的统计特征。与之有一定可比性的是,聚类算法中构造的距离不同会造成样本不同的簇分配结果。欧几里得距离常用于衡量低维空间中任意两点的距离,也可用于衡量具有不同长度的向量之间的距离,是一种通用的距离衡量方法,但应用于高维度的时候会导致计算量随维数呈指数增长的维数灾难(curse of dimensionality)问题。

在采用何种距离进行聚类长度度量的问题上,需要依据聚类对象的特征做出合理选择,以便达到理想的聚类目的。为了对曲线进行聚类,首先需要确定作为研究对象的待聚类曲线的特征适用于哪一种距离量度。时间域电磁法正演响应虽然呈现了连续曲线的形态,实际上是由各时刻测量值为节点的线段首尾相接而成。将其绘制在坐标系中,则 X 轴表示各时刻点,Y 轴表示正演响应值。由于 X 轴取值仅表示时间线上各时刻的贯序序号,因此,单条响应曲线可表示为向量形式:

$$\mathbf{X}=[x_1, x_2, \dots, x_n] \quad (1)$$

式中:  $n$  为时间轴上各时刻;  $x_n$  表示第  $n$  时刻所测量到的正演响应值。由此可见,时间域电磁法正演响应曲线能够以  $n$  维欧几里得空间中具有不同长度的  $n$  维向量的形式表示出来。

### 1.2 距离构建

一种最简单直接的方法是将正演响应曲线间的距离采用多维空间欧几里得距离来表示。欧几里得距离在度量多维空间不同长度向量之间的距离时应用十分广泛。在多维空间中,欧几里得距离为:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (2)$$

式中,  $x_i$  和  $x_j$  分别表示  $n$  维欧几里得空间中以向量形式表示的 2 条不同的曲线,且  $k \in [1, n]$ 。

若同一聚类簇中的各曲线彼此间越紧凑、形状越统一、轮廓越清晰,则该簇的聚类效果越好。然而,采用欧氏距离来衡量曲线间的相似度,仅能保证簇中曲线的紧凑程度较高,却无法对曲线形态相似性做出有效衡量,无法保证聚类后曲线簇的轮廓清晰一致,造成聚类结果中存在偏离中心曲线形态较远的边缘曲线。因此,可以考虑通过加权加以改进,其中权重值用于修正曲线形态的偏离度。加权后的欧式距离为:

$$d(x_i, x_j) = \sqrt{\omega_1(x_{i1} - x_{j1})^2 + \omega_2(x_{i2} - x_{j2})^2 + \dots + \omega_n(x_{in} - x_{jn})^2} \quad (3)$$

式中:  $\omega_1, \omega_2, \dots, \omega_n$  为权重参数;样本向量  $x_i$  与  $x_j$  的加权欧氏距离为  $d_{ij}$ 。

若某响应曲线由  $n$  个测点组成,则构成的  $n$  维欧几里得空间中,2 条不同曲线  $X$  和  $Y$  的  $n$  维向量形式可以表示为:

$$\mathbf{X}=[x_1, x_2, \dots, x_k, \dots, x_n] \quad (4)$$

$$\mathbf{Y}=[y_1, y_2, \dots, y_k, \dots, y_n] \quad (5)$$

式中:  $x_k$  和  $y_k$  分别表示响应曲线  $X$  和  $Y$  上第  $k$  个测点的正演响应值,  $k \in [1, n]$ 。求取曲线向量  $X$  到该簇响应均值曲线向量  $Y$  的加权欧式距离,可依据如下公式:

$$D(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{k=1}^n \omega_k (x_k - y_k)^2} \quad (6)$$

### 1.3 相似性度量

时间域电磁法正演响应曲线可以看做变化趋势连续、轮廓特征明显的曲线。由于地下结构千变万化,不同地下结构产生的正演响应曲线具有不同的形态及距离,或是形态相近距离相近,或是形态相近距离较远,或是距离较近但形态各异,或是形态和距离均有较大的偏差。因此,尽可能将形态和距离均接近的曲线归为一簇,而尽可能在同簇中不要纳入形态或距离与簇均线偏差较大的曲线,是电磁正演曲线聚类需要完成的任务。这就要求对响应曲线聚类的算法在度量曲线相似性时,不仅需要考虑曲线距离,还要考虑曲线的形状、趋势等特征。由于电磁勘探采集的数据是离散数据,将其整理后绘制的正演响应曲线实际上可以看做多个线段组合成的折线,具有各线段端点首尾连接、两端点间连续的特点。

为了将距离和形状纳入聚类相似性度量指标,常见的做法有:(1)根据应用需求对聚类算法中的距离作出不同的定义,如综合曲线形态特征和波动特征,构造曲线双尺度相似性欧式形态距离(党倩等,2022)等;(2)构造加权欧几里得距离,在计算距离时根据实际应用需要进行权重计算,对向量赋予不同权重值;(3)在距离度量中引入表征曲线形状或趋势的参数,如采用标准化的夹角余弦作为样本相关系数来表征曲线形状相似性(朱晓清,2012),采用离散类属性值表达曲线形态变化特征(党倩等,2022)等。加权 K-Means 聚类方法引入序列元素的权重来判断序列元素相似度(吴迪,2014),能够很好地适应时间域电磁法正演响应曲线的聚类需求。

针对时间域电磁法正演响应曲线的特点,根据正演响应曲线的特征定义了权重计算方法,采用加权的欧几里得距离进行曲线相似性度量。加权 K-Means 聚类采用附加了权重系数的加权距离来度量样本相似度,其引入的权重有助于将正演响应曲线的分布特性和形态特性相结合,充分利用曲线携带的特征信息,得到更为准确的聚类结果。

权重参数向量  $\omega_{XY}$  用于衡量曲线间的形态相似程度,可以采用相关系数进行定义:

$$\omega_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{DX} \sqrt{DY}} \quad (7)$$

其中  $X$  和  $Y$  表示  $n$  维欧几里得空间中 2 条不同曲线的向量形式,  $\text{Cov}(X, Y)$  表示  $X$  和  $Y$  的协方差,  $DX$  和  $DY$  分别表示  $X$  和  $Y$  的方差:

$$\text{Cov}(X, Y) = \frac{\sum_{k=1}^n (x_k - EX)(y_k - EY)}{n} \quad (8)$$

$$DX = \frac{\sum_{k=1}^n (x_k - EX)^2}{n} \quad (9)$$

$$DY = \frac{\sum_{k=1}^n (y_k - EY)^2}{n} \quad (10)$$

$EX, EY$  分别表示向量  $X$  和  $Y$  各维度元素的总体均值:

$$EX = \frac{\sum_{k=1}^n x_k}{n} \quad (11)$$

$$EY = \frac{\sum_{k=1}^n y_k}{n} \quad (12)$$

因此,对权重系数  $\omega_k$  的定义为:

$$\omega_k = \frac{\sum_{k=1}^n (x_k - EX)(y_k - EY)}{\sqrt{\sum_{k=1}^n (x_k - EX)^2} \sqrt{\sum_{k=1}^n (y_k - EY)^2}} \quad (13)$$

代入式(4)即可计算得到曲线  $X$  到代表簇中心的簇均值曲线  $Y$  的距离  $D(X, Y)$ 。由此构造完成本文用于曲线聚类的距离的定义及相似性度量方法。

#### 1.4 聚类算法

K-Means 算法是应用广泛的一种基于划分的聚类算法,其主要思路如下:在一组样本中,首先确定  $K$  个中心点,计算样本到达各中心点的加权距离,将加权距离最小的中心点所代表的簇确定为该样本被划归的簇。依次对每个样本进行同样的操作,直到完成所有样本的归簇。然后依据簇中心更新算法对中心点进行修正,修正后由于中心点的变化,各样本与各中心点的距离也产生了变化,需要重新对每个样本执行归簇操作。如此循环,经过多次迭代,最终完成对中心点的修正。每一次迭代都重新计算各样本到达更新后的中心点的距离,然后重新执行归簇操作。直到某一次更新中心点后,各分簇情况达到迭代阈值,满足聚类效果判别算法的最优解,停止迭代。最终得到一个最优的聚类分簇方案。

在聚类过程中,簇中心更新算法为:

$$C_e = \frac{1}{n} \sum_{x_i \in C} x_i \quad (14)$$

式中: $n$  为簇  $C$  中样本个数; $x_i$  为簇  $C$  中的样本。

迭代阈值停止条件为,当样本聚类后划分为  $K$  簇时,每一个样本与所在簇中心的平方距离的总和最小,即:

$$D_{\min}(K) = \sum_{i=1}^K \sum_{x_i \in C_i} \|x_i - \bar{c}_i\|^2 \quad (15)$$

式中: $\bar{c}_i$  为簇  $C_i$  的中心均值线; $x_i$  为簇  $C_i$  中的样本。

聚类的过程可以表示为图 1 所示的示意图。充分利用以向量来表示的曲线所包含的有用信息,将曲线的形状和变化走向等与曲线相似性判别相关的有用信息用权重值的大小体现出来,通过计算时间域电磁法正演曲线各段节点与簇中心均值线的相关系数,作为权重系数附加到各节点距离上再进行求和,以便得到附加了曲线轮廓变化信息的距离,用于聚类归簇判断。

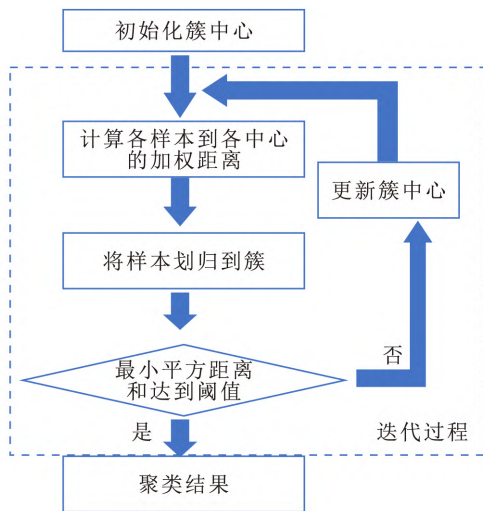


图 1 聚类流程示意图  
Fig. 1 K-Means clustering

### 1.5 效果评价

聚类算法的目的是使得到的结果为全局最优。由于迭代过程存在一定的随机性,如初始中心点的选取等,相同数据和条件下重复同样算法进行聚类得到的结果也可能存在差别<sup>[13]</sup>,为了对聚类的结果有一个科学、客观的评判,需要采用量化的指标来评判聚类效果。聚类效果可以从聚类结果的好坏和聚类数目的优劣 2 个方面进行评估。聚类有效性指标可以分为内部有效性指标和外部有效性指标等,其中内部有效性指标针对聚类结果的同簇聚合程度和簇间分离及差异程度作为评判聚类效果好坏的指标,适应性强,可用于评价无标签数据聚类效果(朱青,2019;党倩等,2022)。常用的聚类内部有效性评价指标有平均指标适应值(MIA)、离散型指标(CDI)、相似矩阵指标(SMI)、紧密度指标(CMP)、分离度指标(PROX)、重叠度指标(CB)、误差平方和总和(SSE)、变异系数(CV)、轮廓系数(SIL)等(朱晓清,2012;吴亚雄等,2020;党倩等,2022)。结合时间域电磁法一维响应曲线的特征,可采用均方差、变异系数来对聚类效果进行评价。

为了评价每次聚类中各簇曲线与簇平均曲线的离散程度,在确定 K 值和聚类中心后,引入均方差(STD)对每簇曲线的聚类效果进行辅助评价:

$$STD = \sqrt{\frac{\sum_{i=1}^n (d_i - d_m)^2}{n - 1}} \quad (16)$$

式中:  $n$  为簇内样本个数;  $d_i$  为簇内第  $i$  个样本与

簇均值之间的距离;  $d_m$  为簇内所有曲线与均值之间的平均距离。

引入变异系数 CV 来衡量聚类前后的变异程度。变异系数为每簇聚类样本与中心线距离的均方差与均值的比值:

$$CV = \frac{STD}{m} = \frac{\sqrt{\frac{\sum_{i=1}^n (d_i - d_m)^2}{n - 1}}}{\frac{\sum_{i=1}^n d_i}{n}} \quad (17)$$

根据前一层聚类结果,计算各簇聚类的均方差和变异系数,并统计每个簇的样本数量,综合考虑是否需要进行下一层聚类。

## 2 地电结构类型划分

地下结构错综复杂,体现在观测数据的响应曲线上则是千变万化的。由于地下电性层的电导率、层厚度、高低阻体形状、异常体分布位置等变量太多且变化太复杂,想要拟合出从响应到地电结构的显式规律十分困难。在实际生产中,我们更关心反演的结果,亦即,通过观测到的电磁响应来推测地下可能性最大的电性结构。而这正是有监督学习所擅长的——通过学习大量已知数据来建立包含隐式映射规律的模型,从而实现将新的未知数据输入模型得到输出的预测数据。由于通过聚类实现了对时间域电磁法正演响应曲线的特征提取,使得进一步对正演曲线样本进行地电结构类型标签化成为可能,进而通过机器学习得到机器对样本集地电结构类型划分的结果。主要步骤如下:

(1)建立时间域电磁法正演响应样本集。有监督学习过程建立在具备 1 个用于模型训练的庞大的先验数据集基础上。对于时间域电磁法而言,这个数据集最理想的情况是由千千万万种地电结构所产生的电磁响应样本组成。时间域电磁法正演响应可以根据某一测点或测线绘制出随时间变化的衰减曲线,也可以根据不同的时间道绘制出随测点位置变化的响应曲线。这些响应曲线通过其外在的形状变化等特征反映了地下结构的内在电性特征。因此,可以通过对模拟生成的地电结构数据进行时间域电磁法正演计算,得到由不同地电结构对应正演响应数据组成的样本集。

(2) 引入聚类算法实现地电结构类型划分,完成时间域电磁法响应数据样本集的标签化。借助无监督学习的聚类算法对样本进行特征提取,完成样本集自动分类,使得每一个样本附上类型标签,建立起标签化后的训练样本集。

(3) 建立从电磁响应到地电结构的映射关系。构建基于有监督学习的多分类任务神经网络,将标签化的训练样本集作为输入,对神经网络进行训练,训练完成后建立起从时间域电磁法正演响应到地电结构的映射模型。

(4) 实现地电结构预测。当得到新的时间域电磁勘探响应数据,将其作为输入数据输入到训练好的映射模型中,以神经网络输出的分类结果作为地电结构类型划分的判定依据,得到预测的地电结构类型。

### 3 应用效果分析

采用一组开源的时间域电磁法观测数据,对论文设计的地电结构类型划分流程进行验证(<https://storage.googleapis.com/simpeg/doc-assets/emldtm.tar.gz>),并将反演结果进行对比。该数据为单测点一维时域数据,接收参数为位于地表上方 20 m,接收线圈是半径为 6 m 的水平环状线圈,接收器测量参数为线圈中心的磁通量垂直分量。发射参数为采用水平环状发射源,发射峰值

的电流幅值为 1A,发射波形为阶跃信号。对时域数据进行反演,得到一维层状电导率模型。

对该组开源数据进行地电结构类型划分。首先由算法随机生成 10 000 个层状地电结构模型组成的样本集,深度为地下 0~250 m,层厚度为 25 m,层电导率根据算法随机生成。根据发射参数和接收参数,采用 SimPEG 开源软件(Cockett et al.,2015)对地电结构模型样本逐个进行单测点时间域电磁法正演,得到正演响应曲线样本集。采用加权 K-Means 聚类算法对正演响应曲线样本集进行聚类,完成样本的标签化。聚类分 2 个层次进行,先根据轮廓系数值的肘部(图 2)确定  $K=5$ ,得到第一层次聚类结果(图 3)及各簇均值线(图 4)。然后进行第二层次聚类,得到 25 种地电结构分类(图 5)及各簇均值线(图 6)。

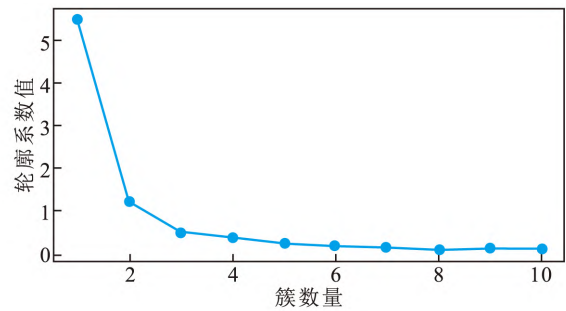


图 2 正演曲线样本集轮廓系数值计算结果  
Fig. 2 Results of calculation of the SSE of the set of forward curve samples

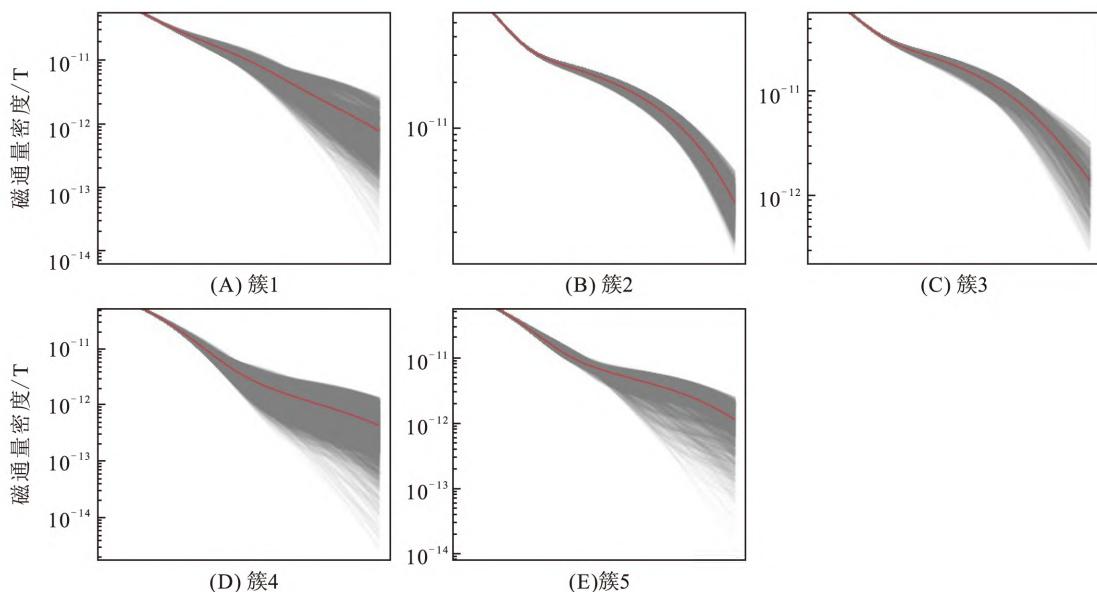


图 3 第一层次聚类结果  
Fig. 3 Results of first-level clustering

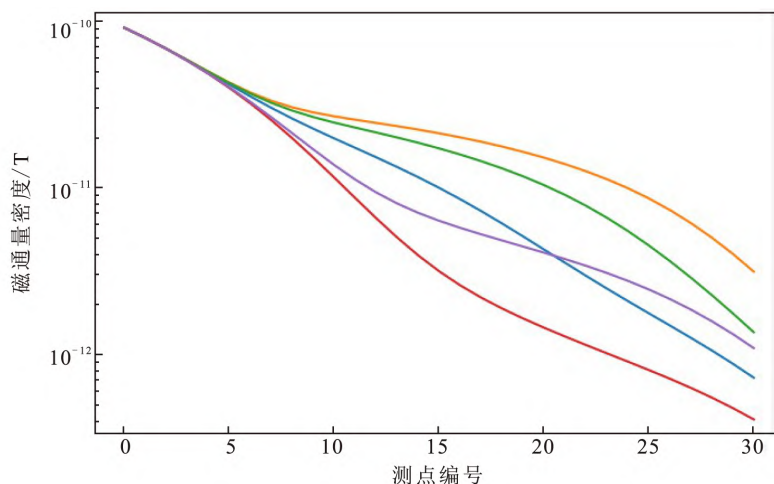


图 4 第一层次聚类结果各簇均值线

Fig. 4 Mean lines of each cluster in the results of first-level clustering

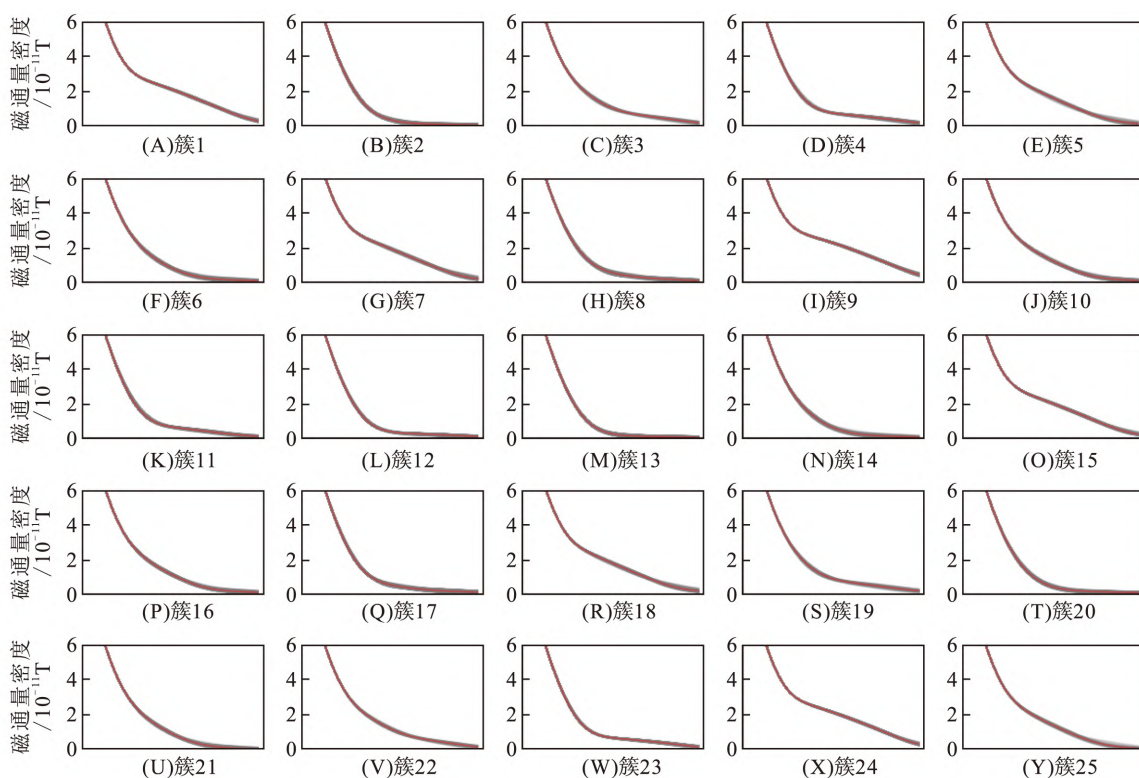


图 5 正演曲线样本集聚类结果

Fig. 5 Results of clustering of the forward curve samples

完成正演曲线样本标签化以后,采用多层感知器(MLP, Multilayer Perceptron)神经网络实现多分类任务。将正演曲线样本集数据输入网络进行训练,训练完成后就建立起了地电结构模型和单点时间域电磁法响应曲线的映射关系。将该组开源的观测数据输入训练好的 MLP 神经网络,映射后得到输出的预测地电结构模型。图 6

展示了实际地电结构模型(黑色)、论文的地电结构类型划分得到的地电结构模型(红色)和 L2 范数正则化反演得到的地电结构模型(蓝色)的对比。图 7 展示了该组观测数据(黑色)、论文方法划分得到的地电结构类型对应的正演响应数据(红色)和 L2 范数正则化反演方法对应的预测响应数据(蓝色)之间的对比。

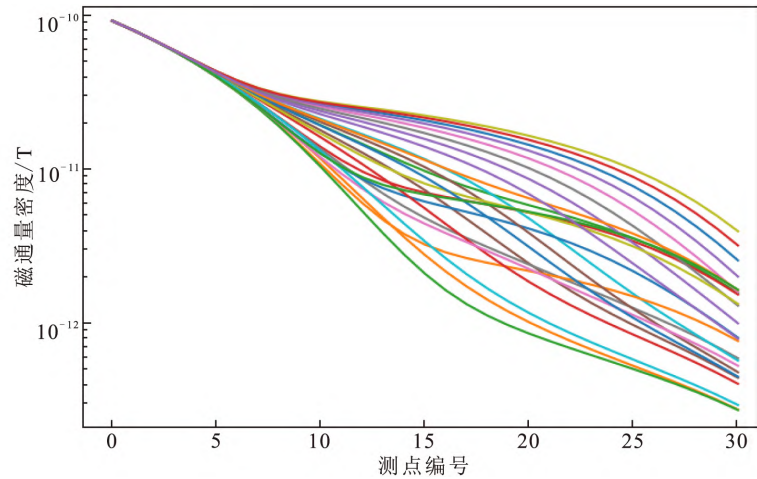


图 6 正演曲线样本集各簇均值线

Fig. 6 Mean lines of each cluster in the sample set of forward curves

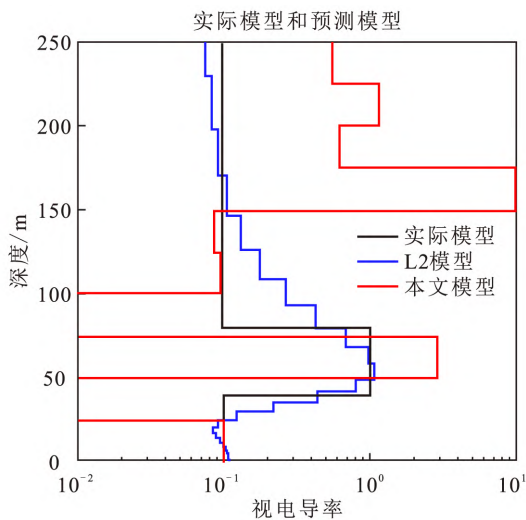


图 7 反演得到的地电结构模型与实际模型的对比

Fig. 7 Comparison between the results of inversion and those of the actual model

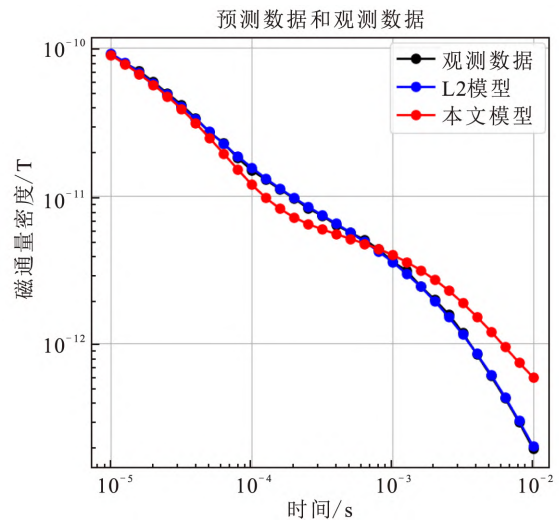


图 8 反演对应的预测响应数据与实际观测数据的对比

Fig. 8 Comparison between the predicted and observed data

由图 7 和图 8 的结果可以看到,地电结构类型划分结果与实际模型相比,对浅层存在的高电导率层、中部存在的低电导率层做出了有效预测,但是对高电导率层以下的连续低电导率层存在错误预测。这是由于采用时间域电磁法进行观测时,无法避免高电导率层的屏蔽作用,从而无法正确反映高电导率层下方的低电导率层。这会导致在采用论文的方法进行模拟数据生成时,不同的地电结构模型样本中由于浅层存在相近位置和相近电导率的高电导率层,而产生相似的正演响应曲线,造成地电结构划分的多解性。该方法虽然有效缩短了地电结构类型预测时间,但预测的准确度仍需要进一步提高,可以考虑的发展方向为

采用更先进的深度学习方法(Sun et al., 2020; Cao et al., 2023)、结合实际地质资料对反演过程进行约束、在生成样本集时通过约束缩小样本生成范围、提高模型与实际情况的契合度并有效提高多分类预测模型的准确度等。

### 4 结论

对模拟生成的不同地电结构模型时间域电磁法正演响应曲线样本进行了分层次的加权 K-Means 聚类,并采用开源数据与其他反演方法进行对比验证分析,结论如下:

- a. 对正演响应曲线进行聚类从而完成地电结构模型类型划分是可行的,聚类结果可以作为类

型标签用于构造多分类任务神经网络的标签化训练样本数据集,以便进一步实现基于深度学习技术的时间域电磁法反演;

b. 目前基于聚类的地电结构预测方法还不能完全替代传统的反演方法,但以此快速得到的预测结果可作为野外施工参考资料;

c. 下一步可考虑结合地质资料约束将该方法进行改进以作为传统反演方法的有效补充。

### [ 参 考 文 献 ]

- 党倩,崔阿军,尚闻博,等,2022. 采用欧式形态距离的负荷曲线近邻传播聚类方法[J]. 西安交通大学学报, 56(1):165-176.
- Dang Q, Cui A J, Shang W B, *et al.*, 2022. Affinity propagation clustering method of typical load curve with Euclidean morphological Distance [J]. Journal of Xi'an Jiaotong University, 56(1):165-176. (in Chinese)
- 胡静,2007. 机器学习及其神经网络分类器优化设计[D]. 合肥:合肥工业大学档案馆.
- Hu J, 2007. Machine Learning and Optimization Design of Neural Network Classifier[D]. Hefei: The Archive of Hefei University of Technology. (in Chinese)
- 黄颖,王雪秋,周子坤,等,2019. 聚类分析算法在大地电磁三维解释中的应用[J]. 地球物理学进展, 34(2): 568-572.
- Huang Y, Wang X Q, Zhou Z K, *et al.*, 2019. Application of clustering analysis algorithm in three-dimensional magnetotelluric interpretation[J]. Progress in Geophysics, 34(2):568-572. (in Chinese)
- 吴迪,2014. 基于加权相似度的序列聚类算法研究[D]. 秦皇岛:燕山大学档案馆.
- Wu D, 2014. Research of Sequence Clustering Algorithm Based on Weighted Similarity[D]. Qinhuangdao: The Archive of Yanshan University. (in Chinese)
- 吴亚雄,高崇,曹华珍,等,2020. 基于灰狼优化聚类算法的日负荷曲线聚类分析[J]. 电力系统保护与控制, 48(6):68-76.
- Wu Y X, Gao C, Cao H Z, *et al.*, 2020. Clustering analysis of daily load curves based on GWO algorithm[J]. Power System Protection and Control, 48(6):68-76. (in Chinese)
- 张贤,李帝铨,李晋,等,2022. 基于特征提取与聚类识别的人工源电磁伪随机信号处理方法[J]. 石油地球物理勘探, 57(4):973-981+1008.
- Zhang X, Li D Q, Li J, *et al.*, 2022. CSEM pseudo-random signal processing method based on feature extraction and clustering identification [J]. Oil Geophysical Prospecting, 57(4):973-981+1008. (in Chinese)
- 赵玄,严家斌,皇祥宇,等,2018. 基于 K 中心点聚类分析的大地电磁阻抗识别[J]. 地球科学与环境学报, 40(6):779-786.
- Zhao X, Yan J B, Huang X Y, *et al.*, 2018. Magnetotelluric impedance recognition based on K-medoids clustering analysis[J]. Journal of Earth Sciences and Environment, 40(6):779-786. (in Chinese)
- 朱青,2019. 基于改进 K-means 算法的民航飞机降落过程油耗分析研究[D]. 合肥:合肥工业大学档案馆.
- Zhu Q, 2019. Analysis and Research on Fuel Consumption of Civil Aviation Aircraft Landing Process Based on the Improved K-Means Algorithm [D]. Hefei: The Archive of Hefei University of Technology. (in Chinese)
- 朱晓清,2012. 电力负荷的分类方法及其应用[D]. 广州:华南理工大学档案馆.
- Zhu X Q, 2012. Load Classification Method and Its Application [D]. Guangzhou: The Archive of South China University of Technology. (in Chinese)
- Cockett R, Kang S, Heagy L J, *et al.*, 2015. SimPEG: An open source framework for simulation and gradient based parameter estimation in geophysical applications[J]. Computers and Geosciences, 85(A):142-154.
- Sun G, Li Y, Yu H F, *et al.*, 2020. Attention distribution guided information transfer networks for recommendation in practice[J]. Applied Soft Computing, 2020, 97:1-13.
- Hartigan J A, Wong M A, 1976. Algorithm AS 136: A k-means clustering algorithm[J]. Journal of the Royal Statistical Society: Series C (Applied Statistics), 1979, 28(1):100-108.
- Kaufman L, Rousseeuw P J, 1990. Finding groups in data: an introduction to cluster analysis [M]. Hoboken: John Wiley & Sons.
- Qiang J K, Luo Y Z, Tang J T, *et al.*, 2010. The algorithm of all-time apparent resistivity for airborne transient electromagnetic (ATEM) survey[J]. Progress in Geophysics, 25(5):1657-1661.
- Song Y C, Meng H D, O'Grady M J, *et al.*, 2010. The application of cluster analysis in geophysical data interpretation[J]. Computers and Geosciences, 14:263-271.
- Cao X J, Sun G, Yu H F, *et al.*, 2023. PerFED-GAN: Personalized federated learning via generative adversarial networks [J]. IEEE Internet of Things Journal, 10(5):3749-3762.

(责任编辑:魏 威)