

徐振, 李兴亮, 龙军. 基于奇异值分解消噪原理的多通道卷积神经网络算法在石油勘探数据质量控制上的研究与实践检验[J]. 智能计算机与应用, 2024, 14(8): 165-168. DOI: 10.20169/j.issn.2095-2163.240827

基于奇异值分解消噪原理的多通道卷积神经网络算法在石油勘探数据质量控制上的研究与实践检验

徐振, 李兴亮, 龙军

(中国石油吐哈油田分公司勘探开发研究院, 新疆 哈密 839000)

摘要: 在石油勘探中, 数据常常受噪声干扰影响, 精度和准确性变差, 数据的可信度降低, 进而影响油气勘探的成效, 因此数据处理是油气勘探中的重要环节。奇异值分解(SVD)是一种常用的信号处理方法, 能有效降低数据噪声。本文提出的基于SVD的多通道卷积神经网络算法, 有针对性地解决石油勘探数据中的噪声问题, 为石油勘探数据的消噪和质量的提升提供了新的思路和方法, 具有良好的推广应用前景。

关键词: 奇异值分解; 多通道卷积神经网络算法; 石油勘探; 数据噪声处理

中图分类号: TE357; TE319

文献标志码: A

文章编号: 2095-2163(2024)08-0165-04

Research and practical verification of multi-channel convolutional neural network algorithm based on singular value decomposition denoising principle in quality control of petroleum exploration data

XU Zhen, LI Xingliang, LONG Jun

(Research Institute of Exploration and Development, PetroChina Tuha Oilfield Branch, Hami 839000, Xinjiang, China)

Abstract: In oil exploration, data is often affected by noise interference, resulting in decreased accuracy and credibility, which in turn affects the effectiveness of oil and gas exploration. Therefore, data processing is an important link in oil and gas exploration. Singular value decomposition (SVD) is a commonly used signal processing method that can effectively reduce data noise. The multi-channel convolutional neural network algorithm based on SVD proposed in this article can better address the problem of noise impact in petroleum exploration data with targeted solutions. This method provides new ideas and methods for noise reduction and quality improvement of petroleum exploration data, and has good prospects for promotion and application.

Key words: singular value decomposition; multi-channel convolutional neural network algorithm; petroleum exploration; data noise processing

0 引言

在当今的石油勘探中, 数据发挥的作用至关重要。但由于受各种噪声因素的影响, 数据的精度和准确性往往损失严重, 数据质量不高、可信度降低, 进而影响油气发现和油气藏的准确预测, 对石油勘探成效影响大。本文提出的基于SVD的多通道卷积神经网络算法, 消除数据处理环节的噪声, 确保数据的准确性和可靠性, 对石油勘探领域中数据噪声处理具有一定的指导意义。

1 基于奇异值分解的消噪策略

1.1 信号消噪原理

奇异值分解是一种常见的数据分解方法, 可用于数据降维、信号处理等应用中。在噪声衰减应用中, 信号数据具有高相关性, 噪声数据具有低相关性。对输入矩阵的奇异值表征数据的相关性进行计算, 得到的奇异值越大, 表示相关性越强^[1]。根据奇异值的大小, 将输入数据分成噪声和信号子空间。

奇异值分解将输入矩阵 X 表示为式(1):

$$X = U\Delta V^T = \sum_{i=1}^R \lambda_i^{\frac{1}{2}} u_i v_i^T \quad (1)$$

其中,正交矩阵 U 由奇异向量 u_i 组成,同时 u_i 也是 XX^T 的奇异向量;正交矩阵 V 由奇异向量 v_i 组成,同时 v_i 也是 $X^T X$ 的奇异向量; Δ 是由奇异值 λ_i 组成的伪对角线矩阵,奇异值 $\lambda_1 \geq \dots \geq \lambda_R \geq 0$, 其排列顺序为单调递减。

如果 $u_i v_i^T$ 是秩为 1 的矩阵,则为数据矩阵 X 的第 i 个特征图像,其 m 和 n 处的样本数据 x_{mn}^i 表示为式(2):

$$x_{mn}^i = u_{im} v_{in} \quad (2)$$

其中, u_{im} 为传播向量, v_{in} 为第 i 个特征图像的时间依赖。

如果数据矩阵 X 为理想纯净信号数据,而且各信号之间是线性独立的,矩阵 X 满秩,恢复信号时需要所有的特征图像,如果各信号之间是线性相关的关系,矩阵 X 的秩为 1,只需要第一个特征图像就可以对信号数据重建^[2]。

输入数据 X 没有噪声时,设矩阵 X 的秩为 M ,在增加噪声后,设矩阵 X 的秩为 N 。假设噪声与信号无关,使用奇异值分解算法,对含有噪声的数据 X 的信号与噪声分离的表达式为式(3):

$$X = X_{\text{信}} + X_{\text{噪}} = \sum_{k=1}^M \lambda_k u_k v_k^T + \sum_{k=M+1}^N \lambda_k u_k v_k^T \quad (3)$$

其中,信号子空间 $X_{\text{信}}$ 由数据 X 的前 M 个特征图表示,噪声子空间 $X_{\text{噪}}$ 由数据 X 的剩余 $N - M$ 个特征图表示。

1.2 复杂输入数据的奇异值分解

使用式(3)能够对单一的噪声干扰数据进行信号和噪声的分离,但是对于多类型的噪声干扰,无法简单地分解为两个子空间进行数据特征表示,也无法进行信号与噪声分离^[3]。需要进一步增加子空间数量,使数据特征得到尽可能的分离,将式(3)改写为式(4):

$$X = \sum_{i=2}^K X_i \quad (4)$$

其中, X_i 表示第 i 个子空间, K 表示子空间个数。

根据式(4),奇异值分解不再只有信号和噪声两个子空间,而是 K 个具有不同特征子空间, K 值与输入数据 X 相关,而且对输出结果起着决定作用。

分别设 K 值为 2、3、4。 K 为 2 时,分解为信号和噪声子空间,此时较多的噪声和信号数据在同一子空间同时存在,数据分解特征不明确; K 为 3 时,分解为高能量信号、剩余信号和噪声子空间,此时数据分解特征具有充足空间; K 为 4 时,信号能量被分解

到 3 个子空间,对信号的结构特征带来破坏,对后续特征提取带来不利影响^[4]。根据计算结果可知,使用奇异值分解为 3 个子空间能够有效地表示数据特征,并且对后续的神经网络算法最有帮助。

综上所述,将式(4)改写为式(5):

$$X = X_{\text{高信}} + X_{\text{剩信}} + X_{\text{噪}} \quad (5)$$

其中, $X_{\text{高信}}$ 表示含有较高奇异值的高能量信号子空间; $X_{\text{剩信}}$ 表示含有剩余较少奇异值的剩余信号子空间; $X_{\text{噪}}$ 表示去除前两种特征的噪声子空间。

2 基于奇异值分解消噪原理的多通道卷积神经网络算法

2.1 基于奇异值分解消噪原理的多通道卷积神经网络结构

基于奇异值分解消噪原理的多通道卷积神经网络结构如图 1 所示。

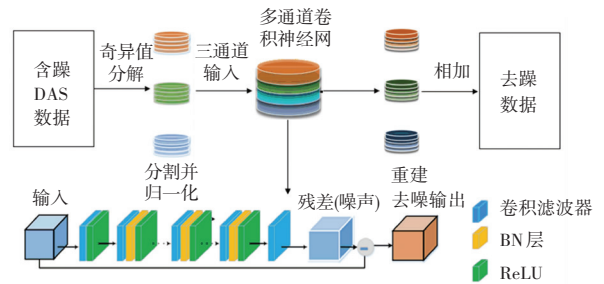


图 1 基于奇异值分解消噪原理的多通道卷积神经网络结构

Fig. 1 Structure of multi-channel convolutional neural network based on singular value decomposition denoising principle

基于奇异值分解消噪原理的多通道卷积神经网络为三通道输入,通过多通道卷积神经网络学习输入的子空间特征,将 3 个子空间特征信息进行去噪映射,三通道输出去噪数据后,将 3 个子空间的去噪数据相加,得到最终去噪结果^[5]。

基于奇异值分解消噪原理的多通道卷积神经网络结构组成:

- (1) 卷积(Conv) + 整流线性单元(ReLU), 提取数据特征并实现非线性映射;
- (2) 卷积(Conv) + 标准化(BN) + 整流线性单元(ReLU), 提高收敛速度,并且不会使梯度消失;
- (3) Conv 对输出数据进行重建。

在基于奇异值分解消噪原理的多通道卷积神经网络结构中,输入数据的大小、网络的层数,以及卷积核的大小等,对于网络输入结果具有直接影响^[6]。为了建立最优的基于奇异值分解消噪原理的多通道卷积神经网络结构,需要通过实验确定网络框架的最优参数。

(1)网络层数的影响实验。设计 10、15、20 和 25 层 4 个不同层数的网络,输入数据的大小、卷积核大小等其他参数不变,卷积层数为 20 层时,训练的损失误差最小,此时加上输入层和输出层,共有 22 层网络层。在卷积层数为 25 层时,观察到训练的损失误差增大,出现网络退化情况,说明网络层过多给特征提取带来影响,选择网络层数为 20 层。

(2)输入数据大小的影响实验。大的数据能够给数据连续特征和相关性带来帮助,但是数据过大,会使计算量增加。设计 32×32 、 64×64 和 128×128 3 种数据大小,训练误差随着数据的增大而减小,在对计算机配置的统筹考虑下,将数据大小设定为 128×128 。

(3)卷积核越大,训练效果越好,但是训练所用的时间和难度也增大^[7]。为使网络训练的效果和效率得到平衡,卷积核大小为 3×3 ,此时第一层的卷积核为 64 个,大小为 $3 \times 3 \times 3$;第二层至二十一层的卷积核为 64 个,大小为 $3 \times 3 \times 192$;输出层的卷积核为 3 个,大小为 $3 \times 3 \times 64$ 。

最终确定基于奇异值分解消噪原理的多通道卷积神经网络结构的最佳参数见表 1。

表 1 基于奇异值分解消噪原理的多通道卷积神经网络结构的最佳参数
Table 1 Optimal parameters for multi-channel convolutional neural network structure based on singular value decomposition denoising principle

网络结构	最佳参数
输入通道	3
网络层数	22
卷积核的大小	3×3 , $3 \times 3 \times 192$, $3 \times 3 \times 64$
Patch 的大小	128×128
Batch 的大小	64×64
学习率	$[10^{-5}, 10^{-3}]$
训练循环	50

为了提升多通道卷积神经网络算法的特征提取能力,并且能够适应数据处理,在网络输入层和输出层增加跳层连接,进行残差学习。残差映射如图 2 所示。

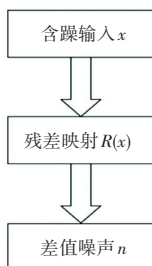


图 2 残差映射

Fig. 2 Residual Mapping

通过残差学习,得到输入映射 $R(X; \theta) \approx N$,其中 $R(\cdot)$ 表示含噪数据 X 和噪声数据 N 之间的映射; θ 表示网络学习参数。消耗结果 S 为含噪数据 X 减去估计噪声数据 $R(X; \theta)$ 。

2.2 基于奇异值分解消噪原理的多通道卷积神经网络训练过程

对深度学习算法的性能带来影响的因素除了网络框架之外,大量而且质量高的训练数据能够使学习精度得到提高。在去噪训练的过程中,需要纯信号和噪声两个数据集,但是在进行实际石油勘探时,没有纯信号数据,影响纯信号数据集的构建^[8]。在基于奇异值分解消噪原理的多通道卷积神经网络训练过程中,使用奇异值分解得到的信号子空间数据和噪声子空间数据分别作为信号样本数据和噪声样本数据,使用残差学习能够使网络集中在对噪声特征的提取,减少了对纯信号数据的依赖。

使用 64×64 的滑动窗截取信号和噪声样本数据,得到信号和噪声训练集样本。在信号和噪声训练集样本中分别随机选择 10 500 个和 8 360 个样本作为信号训练集和噪声训练集。在训练前先遍历信号和噪声训练集,每次迭代随机抽取单个信号和噪声样本,对振幅进行随机缩放,并在信号样本上叠加,最终得到不同信噪比的含噪训练数据^[9]。使用相同步骤建立测试集,共含有 5 634 对训练数据对。

基于构建的网络和训练集,本文提出的基于奇异值分解消噪原理的多通道卷积神经网络算法的消噪流程:

(1)使用奇异值分解得到的 3 个子空间数据作为网络输入,并在训练过程不断将噪声数据向其中加入,得到含噪数据^[10];

(2)使用残差学习的方法,在卷积神经网络中将含噪数据映射为噪声数据;

(3)消噪映射模型训练好后,应用到经奇异值分解后的 3 个子空间数据中;

(4)将 3 个子空间的预测噪声数据输出,分别对 3 个子空间的输入数据与输出数据相减,得到 3 个子空间的数据消噪结果;

(5)将 3 个子空间的数据消噪结果相加,得到最终数据消噪结果。

3 基于奇异值分解消噪原理的多通道卷积神经网络算法在石油勘探数据质量控制上的实践检验

为验证本文所提出的基于奇异值分解消噪原理

的多通道卷积神经网络算法在实际石油勘探数据质量控制应用的有效性,将其应用于实际石油勘探数据进行降噪处理。以某地区实际石油勘探采集数据为例,该数据中包含衰落、水平、振铃、斑状和背景噪声等5种类型噪声干扰,在这5种噪声叠加影响下,实际数据库自治服务(DAS)数据受到严重干扰,甚至对信号轴的结构和位置都难以辨别,传统降噪算法处理该类情况极为困难,数据结果质量很差。

为证明基于奇异值分解降噪原理的多通道卷积神经网络算法在石油勘探数据质量提升上的优势,使用该算法、传统降噪算法和单通道神经网络算法3种算法进行降噪处理。使用传统降噪算法的降噪分辨率低,残留的噪声较多;使用单通道神经网络算法的噪声衰减能力较好;使用基于奇异值分解降噪原理的多通道卷积神经网络算法,不仅能够使多种干扰噪声同时得到消减,而且消噪后得到的数据的保真度、信噪比和分辨率最好^[11]。

对使用3种算法去除的噪声进行对比,传统降噪算法和单通道神经网络算法去除的噪声数据中能够观察到明显的信号轴能量,而基于奇异值分解降噪原理的多通道卷积神经网络算法去除的噪声数据中几乎不存在信号能量泄露的问题。

对3种算法的消噪结果进行区域局部放大对比消噪细节,传统降噪算法的噪声残留比较明显,并且信号轴不清晰;单通道神经网络算法的消噪效果虽然比传统降噪算法有所提升,但还是有噪声残留,并且信号轴能量有衰减;基于奇异值分解降噪原理的多通道卷积神经网络算法能够对噪声有效的抑制,并且能够使信号轴保持完整清晰,具有优越的消噪能力,从而使得数据质量得到较大提升。

对3种算法的信号保幅能力和结果保真度进行对比,传统降噪算法的消噪结果和去除噪声数据之间的相似性很大,消噪能力不强;单通道神经网络算法对于水平噪声的压制不是很彻底,而且存在信号能量泄露问题;基于奇异值分解降噪原理的多通道卷积神经网络算法的信号保幅能力和消噪能力极佳。3种算法得到的结构相似性平均值,基于奇异值分解降噪原理的多通道卷积神经网络算法为 1.025×10^{-3} ,远远小于传统降噪算法的0.019和单通

道神经网络算法的0.004。

综上所述,本文提出的基于奇异值分解降噪原理的多通道卷积神经网络算法具有较好的消噪能力,在石油勘探数据质量控制与提升上具有应用价值。

4 结束语

本文提出基于奇异值分解降噪原理的多通道卷积神经网络算法。经过石油勘探数据质量控制系统的实际数据质量改进的应用实践检验,表明与传统降噪算法和单通道神经网络算法相比,该算法在石油勘探数据质量控制上不仅使多种干扰噪声同时得到消减,而且数据的保真度、信噪比和分辨率最好,在数据噪声处理中有着更强的噪声补偿能力和更精确的局部特征提取能力。基于奇异值分解降噪原理的多通道卷积神经网络算法在石油勘探领域具有重要的应用和推广价值。

参考文献

- [1] 杨德宽,郑泽继,胡立新,等. 对深层地震勘探中随机噪音的一点认识[J]. 石油物探,2000,39(3):13.
- [2] 程发斌,汤宝平,钟佑明. 基于最优Morlet小波和SVD的滤波消噪方法及故障诊断的应用[J]. 振动与冲击,2008(2):91-94.
- [3] 段向阳,王永生,苏永生. 基于奇异值分解的信号特征提取方法研究[J]. 振动与冲击,2009(11):30-33.
- [4] 蔡志东. 井中地震技术:连接多种油气勘探方法的桥梁[J]. 石油地球物理勘探,2021,56(4):922-934.
- [5] 郭淑文,董晓伟,朱伟峰,等. 高分辨率地震处理解释技术在堤北低断阶岩性油气藏勘探中的应用[J]. 录井工程,2020,31(S1):113-119.
- [6] 杨学文,汪如军,邓兴梁,等. 超深断控缝洞型碳酸盐岩油藏注水重力驱油理论探索[J]. 石油勘探与开发,2021,49(1):116-124.
- [7] 李光辉,李月. 沙漠地区地震勘探随机噪声建模及其在噪声压制中的应用[J]. 地球物理学报,2016,59(2):682-692.
- [8] 于四伟,杨午阳,李海山,等. 基于深度学习的地震散射面波智能压制方法[J]. 科学通报,2021,66(18):2343-2354.
- [9] 钟铁,李月. 分布式声传感器井中背景噪声统计特性分析[J]. 吉林大学学报(信息科学版),2021,39(1):8-14.
- [10] 李彦鹏,易维启,曾忠,等. VSP处理新技术及应用[J]. 石油地球物理勘探,2018,53(s2):42-49.
- [11] 李彦鹏,李飞,李建国,等. DAS技术在井中地震勘探的应用[J]. 石油物探,2020(2):242-249.