

基于深度堆栈自编码器的地震相识别

赵云鹤, 硕良勋, 柴变芳, 李增浩

河北地质大学信息工程学院, 河北 石家庄

收稿日期: 2022年9月22日; 录用日期: 2022年10月19日; 发布日期: 2022年10月27日

摘要

地震相是沉积环境分析和储层预测的基础。然而在实际地震勘探中, 标注的地震数据难以获取。针对这些问题, 本文分别借鉴了深度学习与无监督学习的优点, 提出了基于深度堆栈自编码器的地震相识别方法, 该模型利用深度学习更能捕捉地震数据语义信息的优点, 用深度堆栈自编码器对地震数据进行特征提取, 然后用提取的特征进行无监督聚类算法对地震相进行识别。在实际工区的实验结果表明, 该模型相较于直接使用聚类算法识别地震相, 准确率更高。

关键词

地震相识别, 自编码器, 深度学习, 无监督聚类

Deep Stacked Auto-Encoder for Seismic Facies Classification

Yunhe Zhao, Liangxun Shuo, Bianfang Chai, Zenghao Li

School of Information Engineering, Hebei GEO University, Shijiazhuang Hebei

Received: Sep. 22nd, 2022; accepted: Oct. 19th, 2022; published: Oct. 27th, 2022

Abstract

Seismic facies play a key role in depicting subsurface geology horizons and predicting. However, large amounts of annotated seismic data are often unavailable. To overcome these challenges, we proposed a deep stacked auto-encoder network for seismic facies classification, which draws on the advantages of deep learning and unsupervised learning. Deep stacked auto-encoder performs better to exact deep features from seismic data. Then we cluster exacted features to classify seismic facies without the labels for unsupervised learning. The proposed model has been tested successfully on real seismic data fields, which indicates that the model is more effective than clustering directly.

Keywords

Seismic Facies Classification, Auto Encoder, Deep Learning, Unsupervised Clustering

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

地震相是沉积环境分析和储层预测的基础[1], 随着深度学习技术发展, 深度学习辅助地震资料解释也获得了发展。地震相识别可以基于图片分类和语义分割, Dramsch [2]等、Zhao [3]等和 Liu [4]等人借助于图片分类识别地震相, 而闫星宇等[5]、2021年的 Li [6]以及2022年的 Chen [7]等人借助于语义分割技术识别地震相, 前者需要对地震进行切片, 损失上下文信息, 而后者训练所需要的标注数据更多。

这两类深度学习都需要大量标注数据, 然而在实际的地震勘探中, 获取地震相标注数据的难度大、成本高, 且地震数据涉及真实地质数据, 难以数据增强。针对这些问题, 本文提出了基于深度堆栈自编码器的地震相识别方法, 该模型基于自编码器, 无需使用标注数据就能训练, 且可以通过逐层训练, 堆叠多层后获取深层的网络, 深层网络能得到更加抽象的特征, 更能捕捉地震数据的语义信息, 对提取到的特征直接进行聚类, 结果比直接使用聚类的准确率更高, 更符合实际勘探使用。

2. 方法原理

2.1. 自编码器

自编码器[8] (Auto-Encoder, AE)是一种无监督神经网络, 输入输出相等, 也就是将输入拷贝称为输出, 自编码器将原始数据映射到一个特征空间, 改特征空间为隐藏层, 也就是提取特征。所以自编码器由两部分组成, 编码器和解码器, 编码器将每个样本映射到特征空间的编码, 解码器将编码重构为输出。学习目标是 minimized 重构错误(Reconstruction Error, RE):

$$L_{RE} = -\sum_{i \in n} \left\| X^i - g(f(X^i)) \right\|^2 \quad (1)$$

函数 f 和 g 分别代表自编码器解码器和编码器, X 为输入, 共有 n 个输入, f 的输出成为隐藏变量 h 。该隐藏变量可以用来作为后续机器学习模型的输入, 这样经过降维后减少了聚类算法的计算量, 而且无需人工设计提取地震数据特征, 通过自编码器就能得到较于以前聚类更友好的特征。

欠完备自编码器(Under-complete Auto Encoder)的编码器将原始的数据映射到更低维度的特征空间, 学习比原始数据维度要小的隐藏层特征, 这样将会引导自编码器只关注于数据中主要特征, 也避免了恒等变换, 可以起到压缩或者降维的目的。这与主成分分析(Principal Component Analysis, PCA)的功能类似, 但当编码器为深度神经网络时, 激活函数可以为非线性激活函数, 所以自编码器能学习到比 PCA 更强大的非线性特征提取能力。自编码器的解码器则是用隐藏层的特征来重建原始数据。

K 均值聚类(K-means Clustering), 也称为 K-means 聚类, 是以数据或数据特征之间的距离来指导划分, 即两个数据或数据特征之间的距离越小, 则越可能属于同一类。而在具体训练时, 需要根据数据特点和先验知识预先指定好需要划分为几类, 即 K 的值, 然后在这些个数据中随机选 K 个数据作为中心点, 其次根据计算每个数据与中心点的距离将其划分到所属的类别, 最后求得每一类的均值作为新的中心点来进行新一轮的划分。起始的中心点是随机选取的, 所以需要多次重复实验来获得一个较好的结果。

2.2. 深度堆栈自编码器

可以将多个自编码器堆叠到一起, 获取更深层的神经网络, 这样的自编码器称为深度堆栈自编码器 (Deep Stacked Auto Encoder, DSAE)。深层的神经网络提取的特征一般会更加抽象, 更能捕捉数据的语义信息。而堆栈自编码器使用逐层训练 (Layer-Wise Training) 方法更新参数, 如图 1 所示: 每一层的目标函数均为公式 1 的 MSE 损失, 学习算法就是前面的逐层训练, 先学习第一层自编码器, 然后用隐藏变量学习第二层自编码器以此类推, 直到最终的隐藏变量。

本文每一层的自编码器都更通用化的卷积神经网络, 卷积层、BN 层和 ReLU 层。其中地震切片的通道数变化为 1 到 32, 32 到 64, 64 到 128。然后拉伸后经过全连接层, 获取到长度为聚类个数的特征向量。

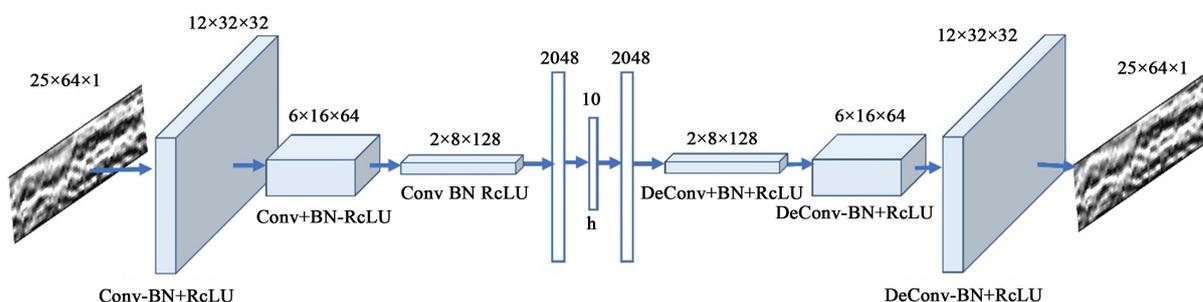


Figure 1. Deep stacked auto-encoder

图 1. 深度堆栈自编码器

3. 实验和结果

3.1. 地震数据介绍

本文用到的数据是荷兰北海 F3 的叠后地震数据集, 该数据集包含 384 平方千米的地震数据, 分别有 951 条主测线和 651 条联络测线, 每个地震道有 462 个采样点, 采样间隔为 4 ms, 所以三维地震体的大小为 $951 \times 651 \times 462$ 。标注数据来源于文献[9], 提供了 25×64 的地震子剖面 (tiles) 和相应的类别, 如图 2 所示, 从上到下十类不同的地震相, 主测线和联络测线分别为 94400 和 94720 个地震子剖面。

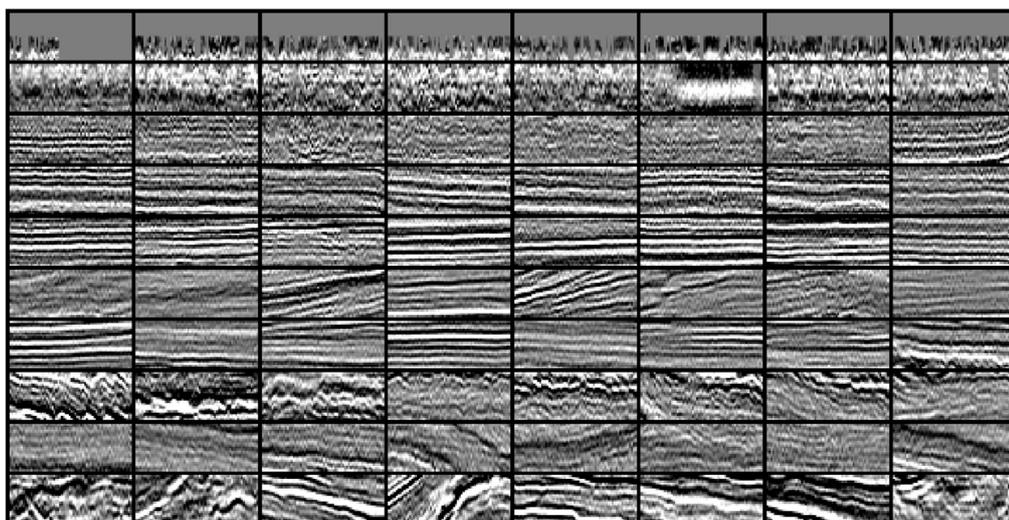


Figure 2. Ten different seismic facies

图 2. 十类地震相

3.2. 模型和评价指标

本文使用的模型具体结构如图 1 所示。地震子剖面经过每一层使用逐层训练进行特征提取：因为共有十类地震相，所以本文采用的 25×64 的地震剖面的隐藏维度为 10。本文实验是用得到的特征向量使用 K 均值算法进行聚类，并与 25×64 的地震剖面直接使用 K 均值算法的聚类结果进行比较。实验结果评价指标采用准确率(Accuracy, ACC)，即正确预测的结果的样本数与总样本数之比，对模型效果的进行比较和评价。

3.3. 实验结果与结果分析

按数据集分为三种，分别是联络测线、主测线上的地震子剖面以及二者所有的地震子剖面，三个数据集的地震子剖面数量分别为 94400、94720、189120。实验结果如表 1 所示：

Table 1. Resulting data of experiment

表 1. 实验结果数据

数据集	K 均值准确率/%	DASE 准确率/%
主测线	36.5	44.6
联络测线	33.6	40.5
所有数据	34.4	45.2

从表中的结果准确率可以看出，用深度堆栈自编码器提取的特征进行 K 均值聚类，效果更好。在单独的主要测线和联络测线的准确率分别提高了 8.1 和 6.9 个百分点，说明模型在主测线上的效果更好一些。而整体数据的准确率直接提升了 9.8%，说明数据越多更能从地震数据挖掘出内在先验知识，从而提高了聚类的效果。

4. 结论

针对地震数据标注成本高、工作量大的问题，本文提出了一种基于深度堆栈自编码器的地震相识别模型。该模型借鉴深度学习和无监督学习的优点，使用深度堆栈自编码器来提取地震数据的低维特征，提取到的特征不仅是聚类友好的而且压缩维度后能有效减少聚类的计算量。将此算法应用于实际工区数据 F3 工区，并在实际工区数据中对提取的特征使用 K 均值算法得到的结果和直接使用 K 均值算法进行比较，发现利用该模型提取特征聚类效果更好，该特征是聚类友好的。

基金项目

河北省教育厅重点项目(ZD2020175)；河北省高校基本科研业务费资助，河北地质大学科技创新团队项目资助(KJCXTD-2021-11)；河北省重点研发计划项目(项目名称：基于时空大数据及深度学习的地质灾害风险识别关键技术研究编号：22375415D)。

参考文献

- [1] 吴宝年. 油藏模型含油饱和度变化的地震振幅响应特征[J]. 物探与化探, 2015, 39(6): 1271-1277.
- [2] Dramsch, J.S. and Luthje, M. (2018) Deep-Learning Seismic Facies on State-of-the-Art CNN Architectures. *SEG Technical Program Expanded Abstracts*, Anaheim, 17 October, 2018, 2036-2040. <https://doi.org/10.1190/segam2018-2996783.1>
- [3] Zhao, T. (2018) Seismic Facies Classification Using Different Deep Convolutional Neural Networks: 88th Annual In-

ternational Meeting. *SEG Technical Program Expanded Abstracts*, Anaheim, 17 October, 2018, 2046-2050.
<https://doi.org/10.1190/segam2018-2997085.1>

- [4] Liu, M., Jervis, M., Li, W., *et al.* (2020) Seismic Facies Classification Using Supervised Convolutional Neural Network and Semi-Supervised Generative Adversarial Networks. *Geophysics*, **85**, 1-57.
<https://doi.org/10.1190/geo2018-0470.1>
- [5] 闫星宇, 顾汉明, 罗红梅, 等. 基于改进深度学习方法的地震相智能识别[J]. 石油地球物理勘探, 2020, 55(6): 1169-1177.
- [6] Li, F.Y., Zhou, H.L., Wang, Z.Y. and Wu, X.M. (2021) ADDCNN: An Attention-Based Deep Dilated Convolutional Neural Network for Seismic Facies Analysis with Interpretable Spatial-Spectral Maps. *IEEE Transactions on Geoscience and Remote Sensing*, **59**, 1733-1744. <https://doi.org/10.1109/TGRS.2020.2999365>
- [7] Chen, X.Y., Zou, Q., Xu, X.X. and Wang, N. (2022) A Stronger Baseline for Seismic Facies Classification with Less Data. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, Article No. 5914910.
<https://doi.org/10.1109/TGRS.2022.3171694>
- [8] MacQueen, J.B. (1967) Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. Berkeley, 281-297.
- [9] Silva, R.M., Baroni, L., Ferreira, R.S., *et al.* (2019) Netherlands Dataset: A New Public Dataset for Machine Learning in Seismic Interpretation. <https://arxiv.org/abs/1904.00770v1>