

基于傅里叶功率谱样本熵和机器学习算法的地震判别与震级预测研究

Yuan Gao, Shuang Xu, Fei Wang*

牡丹江师范学院 黑龙江牡丹江

【摘要】本文主要通过加窗傅里叶变换和功率谱样本熵提取地震波数据的特征，并分别建立支持向量机分类模型和随机森林回归模型，对地震事件进行分类和地震等级预测。结果表明，对于地震识别，测试集的准确率达到 82.7%；对于等级预测，MAE 达到 0.58。因此，分类支持向量机和回归随机森林可以用于地震识别和等级预测。此外，波形信号的功率谱样本熵可以用来衡量波形的特征。基于傅里叶功率谱样本熵和机器学习算法的地震识别和震级预测方法在地震监测预警领域具有潜在的应用价值。这些研究成果为地震相关决策提供了可行的技术工具，有助于提高地震灾害的预测和响应能力。但仍需进一步的研究和验证，以进一步改进和优化方法的性能和稳定性。

【关键词】傅里叶变换；样本熵；支持向量机；随机森林

【收稿日期】2025 年 5 月 3 日

【出刊日期】2025 年 6 月 10 日

【DOI】10.12208/j.ghrm.20250004

Research on earthquake discrimination and magnitude prediction based on Fourier power spectrum sample entropy and machine learning algorithm

Yuan Gao, Shuang Xu, Fei Wang*

Mudanjiang Normal University, Mudanjiang, Heilongjiang

【Abstract】In this paper, the characteristics of seismic wave data are extracted mainly through windowed Fourier transform and power spectrum sample entropy. A support vector machine classification model and random forest regression model are respectively established to classify seismic events and predict the grade of the earthquake. The results showed that for earthquake discrimination, the accuracy of the test set reached 82.7%; For grade prediction, the MAE reached 0.58. Therefore, classification support vector machine and regression random forest can be used for earthquake identification and grade prediction. Additionally, the power spectrum sample entropy of the waveform signal can be used to measure the characteristics of the waveform. The earthquake discrimination and magnitude prediction methods based on Fourier power spectrum sample entropy and machine learning algorithms have potential applications in the field of earthquake monitoring and early warning. These research results provide a feasible technical tool for earthquake-related decision-making and help improve the prediction and response capability of earthquake hazards. However, further research and validation are still needed to further improve and optimize the performance and stability of the method.

【Keywords】Fourier transform; Sample entropy; Support vector machine; Random forest

1 简介

地震是一种较为复杂的地壳运动现象，全球每年都会发生无数次地震灾害。然而，随着城市建设工程项目的快速增加和地震台网监测规模的扩大，一些非自然地震事件如爆破地震、矿震、武器试验、坍塌等发生，

干扰了地震事件的准确预报。在地震监测中有效识别自然地震事件，消除异常干扰信号，对地震预警预报技术、减轻地震灾害具有重要的现实意义。震级预测也是地震预报的重要目标之一，地震级别的准确判定依赖于对大量历史事件的特征挖掘和地震波能量的估算，

*通讯作者：Fei Wang

注：本文于 2023 年发表在 Journal of Applied Mathematics and Computation 期刊 7 卷 3 期，此文为其授权翻译版本。

这有助于制定有针对性的地震应急预案,减少灾害损失^[1]。

本文将利用窗口傅里叶变换和功率谱样本熵的方法,根据仪器记录的地震波数据,提取波的特征变量,建立向量机分类模型和随机森林回归模型,对地震波和非地震波进行分类,并进行地震等级预测,希望为解决地震分类和等级预测问题提供参考。

2 基于功率谱样本熵的波浪特征提取

地震信号数据信号为离散波形信号,本次选取天府第三届全国大学生数学建模竞赛 A 题的数据,其中

包含 120 条天然地震波数据和 30 条非天然地震波数据,并且所有 120 条地震数据均标注了地震级别。

通过观察天然地震波和人工地震波数据的波形图发现,天然地震波波形包络变化缓慢,逐渐增长并回落;而对于爆破等产生的非天然地震波,波形包络在起始位置发生突变,并在短时间内从零突然增长到一个很大的值,如图 1 所示。因此,利用波形信号处理工具傅里叶变换提取功率谱,并用样本熵来度量功率谱分布和变换,从波形数据中提取特征^[2]。

2.1 添加窗口

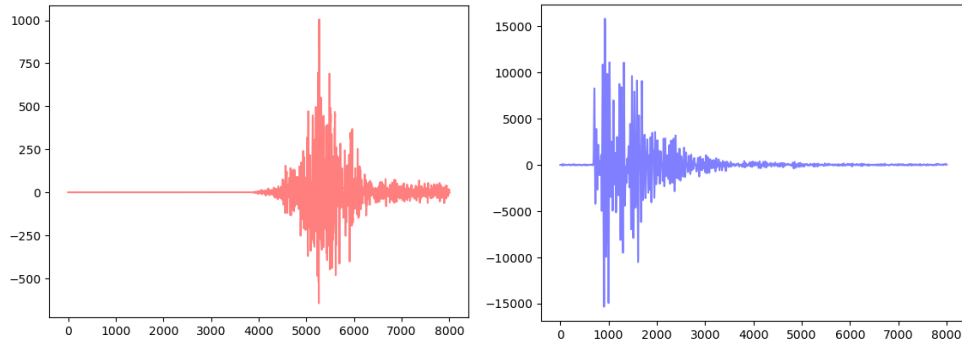


图 1 特征提取流程图

加窗的目的是提取时间序列片段。如果直接截断数据,进行傅里叶变换时会出现谱泄漏,因此顾氏采用加窗处理来减少谱泄漏。本次采用的是汉明窗:

$$\omega(n) = a_0 - (1 - a_0) \cdot \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad (1)$$

其中 $a_0 = 0.53836$, 窗口长度 N 以被分析数据的基波周期为基准,采用略大于基波周期的 2 的整数幂作为窗口长度。

2.2 傅里叶变换

傅里叶变换可以将信号从时间域变换到频率域,得到波形信号各频率分量的幅度和相位信息,便于分析信号特性^[3]。由于电流数据是离散的,因此需要利用离散傅里叶变换,而信号的正逆变换公式 $x(n)$ 其光谱 $X(k)$ 为:

$$X(k) = DFT(x(n)) = \sum_{n=0}^{N-1} x(n) W_N^{kn}, k = 0, 1, \dots, N-1 \quad (2)$$

$$x(n) = IDFT(X(k)) = \sum_{k=0}^{N-1} X(k) W_N^{-kn}, n = 0, 1, \dots, N-1 \quad (3)$$

处理中采用快速傅里叶变换算法,利用 W_N^{kn} 卷积周期 N 与加窗长度相同的周期性,提高计算效率。

2.3 计算功率谱

得到频谱变化图后,选取 Mayer 滤波器组进行频域分割,对频带进行对数能量统计,得到每个窗口帧随时间变化的对数功率谱,通过公式计算:

$$s(m) = \ln\left(\sum_{k=0}^{N-1} |X_a(k)|^2 H_m(k)\right), 0 \leq m \leq M \quad (4)$$

其中 M 为频带数,这里选择 26, $H_m(k)$ 为带内滤波函数, N 为频带内频点数。spectrum 的功率谱计算结果示例 $X_a(k)$ 如图 2 所示。

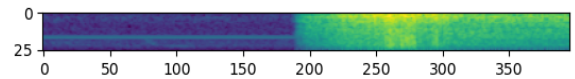


图 2 分段功率谱

2.4 计算频带样本熵

样本熵用于度量序列的分布复杂度。地震数据和非地震数据的分布有明显差异,地震数据初始分布平缓,而非地震数据初始生成于一个突变的位置,因此需要样本熵来度量功率数据在时间序列上的分布复杂度,以此来识别地震类别,预测地震级别。

向量序列的样本熵 $X_m(i)$ 计算如下:

(1) 定义距离函数

$$d[X_m(i), X_m(j)] = \max_{k=0, \dots, N-1} (|x(i+k) - x(j+k)|) \quad (5)$$

并设定距离阈值 r 。计算与距离小于 r 的向量个数 $X_m(i)$ ，记为

(2) 定义：

$$B_i^m(r) = \frac{1}{N-m-1} B_i \quad (6)$$

并计算所有 i 的值。然后计算

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \quad (7)$$

(3) 将 m 的值加 1，得到：

$$A^m(r) = B^{m+1}(r) \quad (8)$$

(4) 计算样本熵：

$$SampEn(m, r, N) = -\ln \frac{A^m(r)}{B^m(r)} \quad (9)$$

这里选取 $m=3$ 、 $r=1$ ，计算每条地震波形数据功率谱的样本熵，并在样本熵的后一列附加是否为天然地震的标签，构建分类数据，若附加震级数据，则构建震级回归数据，选取 10 条分类数据。

3 构建机器学习分类和回归模型

在处理分类和回归等任务时，目前经典有效的方法主要是机器学习分类和回归模型^[4]。机器学习算法可以从数据中学习数据蕴含的内在规律，从而构建相应的数学模型来表征相应特征与目标结果之间的关系，从而实现相应任务的求解。针对真假地震分类预测和地震级别回归预测，本文主要采用支持向量机分类模型和随机森林回归模型。

3.1 支持向量机分类模型

支持向量机模型是一种二分类模型，其基本思想是找到将数据集正确分类为二。

其基本思想是找到能够将数据集正确划分为两个几何间隔的最大值的超平面。简单地说，用一条线（ n 维的超平面）将训练数据集分成两类，当添加新数据时，可以通过这条线对数据进行分类。在支持向量机模型中，求解最优决策边界线（超平面）的问题通常转化为求解两类数据集之间最大间隔的问题，然后将该间隔的中心作为此模型的决策边界：

$$\arg \max_{w, b} \frac{2}{\|w\|} \quad (10)$$

$$s.t. y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, m \quad (11)$$

当要处理的问题是线性不可分的分类问题时，原维度下的决策超平面无法正确解释数据^[5]。通过空间映射变换函数将原空间下线性不可分的数据集在空间上变换为线性可分的，在新的维度空间中再次求解决策

超平面。

映射变换函数使得求内积变得困难，因此可以用核函数代替原维度

直接计算向量点积。常用的核函数有多项式核、径向基函数核等。本例中，选择最高子项为 3 的线性核函数。

3.2 随机森林回归模型

随机森林是一种以决策树模型为基础，采用并行集成学习方法（bagging）构建的弱分类模型。

决策树回归是一种与分类决策树类似的弱分类器，主要包含 CART（分类和回归树）算法的计算过程，主要是根据损失评估决策寻找最优割点形成分支节点，然后针对每个分支区域递归求解该过程，最终形成具有树形结构的回归树。

依靠集成学习强大的泛化能力，随机森林对相应任务有很好的处理能力^[6]。随机森林回归模型中决策树的生长过程是随机的，随机从训练集中选择样本，并从这些样本中随机选择特征进行训练，从而导致不同的训练结果。随机森林回归模型的优势在于，由于最初无法得知各个特征对结果的影响大小，因此采用随机过程可以减小其对分类结果的影响，随机森林回归模型的最终决策结果由每棵决策树的投票决定。

这里选取决策树的数量 100，树的最大深度为 10，最大叶节点数为 50，节点分类损失评估策略采用 MSE。

4 实验结果与分析

实验利用从数据集中提取的功率谱样本熵数据对训练数据和测试数据进行切片，切片比例为 7:3，分别用相应的数据训练支持向量机分类模型和随机森林回归模型，然后用测试数据对得到的模型进行测试。

对支持向量机分类模型进行了测试，测试结果如表 1 所示，训练集数据和测试集数据对模型的准确率均达到了 82% 以上，表明所得模型更好地反映了原始数据的特征，具有良好的泛化和预测能力。

再次对随机森林回归模型进行检验，选取 15 次检验的结果绘制在图 3 中，模型的评估结果如表 2 所示。从预测结果图中可以看出，模型对大部分数据都能做出很好的预测，测试集的 MAE 为 0.589，说明预测误差的平均绝对值约为 0.6，大约是半个震级多一点，说明模型对地震震级有很好的预测能力。

5 结论

从实验结果来看，所选的两个支持向量机和随机

表 1 分类模型评估结果

	准确率	召回率	准确率	F1
训练集	0.822	0.822	0.676	0.742
测试集	0.827	0.827	0.684	0.749

表 2 随机森林回归模型评估结果

	均方误差	均方根误差	平均辅助能量吸	甲基丙烯酸酯共聚物	R ²
训练集	0.104	0.323	0.261	4.353	0.929
测试集	0.601	0.776	0.589	9.701	0.652

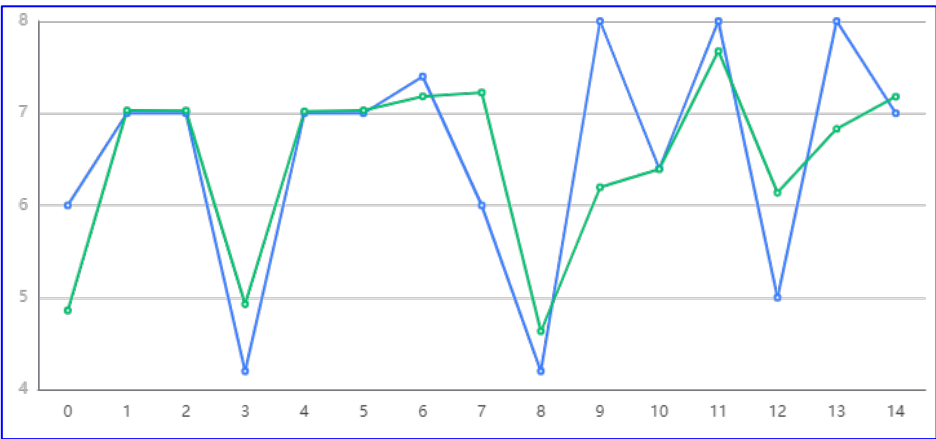


图 3 随机森林回归模型检验结果

森林机器学习模型在相应任务上均能取得良好的结果指标,可用于地震判别和震级预测,同时也表明对于地震波形信号,功率谱样本熵能够很好地表征地震波形。然而,从随机森林回归模型的测试结果可以看出,测试集的性能比训练集差,存在过拟合现象。下一步可以进一步处理过拟合问题,例如进行正则化,或者提高数据集的数量和质量,以减轻或消除过拟合问题。

6 讨论

基于傅里叶功率谱、样本熵和机器学习算法的地震识别与震级预测是一个很有前景的研究领域,未来可以从以下方面进一步研究:

改进特征提取方法:傅里叶功率谱和样本熵是地震信号分析中常用的特征,但仍有改进空间。可以探索使用其他谱分析方法,例如小波变换和时频分析,以获取更准确、更全面的特征信息。此外,还可以结合震源机制和地壳结构等其他地震学参数,构建更复杂的特征表示。

引入深度学习算法:传统机器学习算法在地震判

别和震级预测方面取得了一定的成果,但深度学习算法拥有更强大的表征能力和特征学习能力。未来的研究可以探索利用深度学习算法,例如卷积神经网络、循环神经网络和注意力机制,来提升地震判别和震级预测的准确性和鲁棒性。

数据集和样本标记:地震数据采集和标记对于研究的重要性怎么强调也不为过。未来的研究可以收集更多地震数据,并更细致、更准确地标记数据。同时,还可以考虑对多种地震事件类型(例如地壳震动、地下水动力学和地下岩石运动)进行分类的问题,以及预测不同地震震级的问题。

多模态数据融合:地震研究除了地震信号数据外,还可以结合其他相关的多模态数据,例如地表形变数据、地磁数据、地震波形数据等,进行综合分析和预测。通过融合多种数据源的信息,可以提高地震识别和震级预测的准确性和可靠性。

实时监测预警:地震的实时监测预警对于减少地震灾害具有重要意义。未来的研究可以将基于傅里叶

功率谱样本熵的地震判别与震级预测以及机器学习算法应用于实时监测系统, 实现快速准确的地震识别与预警。

综上所述, 基于傅里叶功率谱样本熵和机器学习算法的地震判别与震级预测研究具有广阔的发展前景。该研究具有广阔的发展前景。未来的研究可以对特征提取、算法模型、数据集与样本标注、多模态数据融合等方面进行深入探索, 以提高地震监测预报的精度和可靠性, 为地震灾害防治提供更有效的支持。

参考文献

- [1] Renouard, A., Maggi, A., Grunberg, M., Doubre, C., & Hibert, C. (2021). Toward false event detection and quarry blast versus earthquake discrimination in an operational setting using semiautomated machine learning. *Seismological Research Letters*, 92(6), 3725-3742.
- [2] Yang, D. H., Zhou, X., Wang, X. Y., & Huang, J. P. (2021). Micro-earthquake source depth detection using machine learning techniques. *Information Sciences*, 544, 325-342.
- [3] Saini, K., Kalra, S., & Sood, S. K. (2022). An Integrated Framework for Smart Earthquake Prediction: IoT, Fog, and Cloud Computing. *Journal of Grid Computing*, 20(2), 17.
- [4] Xiong, P., Marchetti, D., De Santis, A., Zhang, X., & Shen, X. (2021). SafeNet: SwArm for earthquake perturbations identification using deep learning networks. *Remote Sensing*, 13(24), 5033.
- [5] Toh, G., & Park, J. (2020). Review of vibration-based structural health monitoring using deep learning. *Applied Sciences*, 10(5), 1680.
- [6] Mukherjee, T., Singh, C., & Biswas, P. K. (2021). A Novel Approach for Earthquake Early Warning System Design using Deep Learning Techniques. *arXiv preprint arXiv:2101.06517*.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

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



OPEN ACCESS