

基于多台输入的图注意力网络震级估计研究

余仲黎^{1,2}, 朱景宝^{1,2}, 李山有^{1,2}, 宋晋东^{1,2}

(1. 中国地震局工程力学研究所 地震工程与工程振动重点实验室, 黑龙江 哈尔滨 150080;
2. 地震灾害防治应急管理部重点实验室, 黑龙江 哈尔滨 150080)

摘要: 震级估计是地震预警中的重要任务之一。准确的震级估计对于地震影响范围的快速判断和地震预警信息的及时发布至关重要。现有的方法通常基于单个台站的加速度时程提取特征信息进行震级估计, 再通过多台平均的方法得到结果。文中利用多台输入的图注意力网络算法构建端到端的震级估计模型(GAT_M), 作为GAT_M模型输入的是首台P波触发后3s内的多台站地震加速度时程。本研究利用日本防灾科学技术研究所K-NET强震观测台网的强震资料进行模型训练和测试实验。研究表明: 在首台P波触发后3s, 震级估计的平均误差和标准差分别为-0.077和0.40, R^2 为0.72。本研究还分析了震级、时间窗和台站数量对GAT_M模型性能的影响。同时, 在首台P波触发后3s, 与传统Pd方法相比, GAT_M模型有更小的震级估计误差, 在复杂样本数据的情况下, GAT_M模型有较大的优势且能够更好地应用于地震预警震级估计中。

关键词: 图注意力网络; 地震监测预警; 震级; 多台输入

中图分类号: P315.3; P315.7; P315.92

文献标志码: A

Magnitude estimation of graph attention networks based on multi-station inputs

YU Zhongli^{1,2}, ZHU Jingbao^{1,2}, LI Shanyou^{1,2}, SONG Jindong^{1,2}

(1. Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150080, China; 2. Key Laboratory of Earthquake Disaster Mitigation, Ministry of Emergency Management, Harbin 150080, China)

Abstract: Earthquake magnitude estimation is one of the important tasks in earthquake early warning. Accurate earthquake magnitude estimation is critical to quick judgment of earthquake influence areas and timely release of earthquake warning information. Existing methods usually extract the characteristic information based on the acceleration time history of a single station to estimate the magnitude, and then obtain the result by the multi-station averaging method. In this paper, an end-to-end magnitude estimation model (GAT_M) is constructed using a multi-input graph attention network algorithm. The time history of multi-station seismic acceleration within 3 s after the first P-wave is triggered is input into the GAT_M model. The multi-station seismic acceleration waveforms within 3 s after the first P-wave are used as the input of the GAT_M model. In this study, the strong earthquake data from of the K-NET strong earthquake observation network of Japan Institute of Disaster Prevention Science and Technology were used for model training and test experiments. Within 3 s after the first P-wave triggers, the mean error and standard deviation of magnitude estimation are -0.077 and 0.40 respectively, and R^2 is 0.72. The effects of magnitude, time window and number of stations on the performance of GAT_M model are also analyzed. Simultaneously, within 3 s after the initial P-wave triggers, the GAT_M model demonstrates a reduced magnitude

收稿日期: 2024-03-04; 修回日期: 2024-04-07

基金项目: 中国地震局工程力学研究所基本科研业务费专项(2024C05); 国家自然科学基金项目(42304075, 51408564)

作者简介: 余仲黎(1998—), 男, 硕士研究生, 主要从事地震预警研究。E-mail: 542347624yu@gmail.com

通信作者: 宋晋东(1980—), 男, 研究员, 博士, 主要从事地震预警与人工智能地震学研究。E-mail: jdsong@iem.ac.cn

estimation error compared to the traditional Pd method. In the case of complex sample data, the GAT_M model has a greater advantage and can be better applied to magnitude estimation.

Key words: graph attention network; earthquake monitoring and early warning; magnitude; multi-station

0 引言

地震预警系统是一种在地震波到达之前提供几秒到几十秒警报的技术,主要基于地震 P 波传播速度大于破坏 S 波和面波,以及电磁波传播速度远远大于地震波波速的原理,通过监测站捕捉 P 波并迅速分析,然后在 S 波到达并造成破坏之前发出警报^[1-2]。到目前为止,地震预警系统已经在多个国家和地区进行了部署和应用,如日本、墨西哥、美国、意大利、中国等,并取得了较好的减灾效果^[3-12]。

震级估计作为地震预警系统的关键环节之一,对于分析地震影响范围和地震信息发布具有至关重要的作用。现有震级估计方法主要是利用从单个台站记录的加速度时程中提取的幅值信息和频率信息来确定震级大小,然后利用多台平均来确定最终震级^[3-7]。峰值位移(peak displacement, Pd)估计方法是目前广泛应用于地震预警系统中的震级估计方法,主要利用 P 波的幅值信息(一般为 P 波的峰值位移)及地震震源距离与震级之间的关系来估算震级。这种方法可以在地震发生后的几秒到几十秒内迅速提供震级估计,对地震的快速响应和预警是非常重要的。然而,依赖人工提取的有限的幅值信息可能并不足以表征实际震级,并且由于源辐射模式、场地条件和异常噪声等影响因素的存在,可能会导致不同台站的结果存在很大差异^[8-12]。近年来,随着计算能力的提升和人工智能技术的发展,深度学习技术逐渐被用于解决各种地震问题,如地震震相拾取^[13-14]、地震定位^[15-16]和地震预警^[17-24]等。深度学习方法能够通过训练大量数据提高模型的精度,而高密度的地震台网记录也给深度学习方法提供了训练的可能性。许多学者已经开始使用深度学习方法来解决震级估计问题。基于单个台站的加速度时程数据输入,MOUSAVI 等^[25]设计了一个由卷积神经网络和递归神经网络组成的回归模型用于估计地震震级,取得了较好的效果。KUANG 等^[26]提出基于全连接卷积神经网络的震级确定模型 MagNet,通过提取多台站加速度时程数据的全局特征进行地震震级预测,同时, MagNet 限制了输入台站数量为 20 个,并且作为模型输入的加速度时程长度达到了 80 s,这使该方法在实际地震预警应用方面受到限制。

卷积神经网络非常适合欧几里得数据结构,可以从一维信号或二维图像中提取出深度特征信息,但是当面对震源距不同的多台站地震数据时效果并不好。因此本文引入图数据结构对多台站地震数据进行重新构造。图数据结构的一个重要特征是,它们不是由数据的顺序定义的,而是由数据之间的关系定义的。用图数据结构来表征地震台站网络可以更好地利用台站之间的空间关系信息,从而达到更好的效果。然而,对图的有效操作需要保持数据顺序不变^[27]。因此,卷积神经网络无法直接在图上定义卷积。而图神经网络(graph neural network, GNN)作为一种专门处理图数据结构的深度学习工具,近年来获得了许多研究者们的青睐。图神经网络是在图论和深度学习的基础上发展起来的,目前已广泛应用于社交网络^[28-29]、自然语言处理^[30-31]和地震数据分析^[32-33]等领域。VAN DEN ENDE 等^[32]提出了一种基于深度 GNN 的方法来自动化地震源特征分析。该方法通过处理多台站加速度时程数据并明确地纳入地震台站的地理位置信息来估计地震的位置和震级。MCBREARTY 等^[33]通过构建地震参数数据集,然后使用图卷积神经网络进行训练,以用来估计地震位置和震级。以上研究将地震信号转换为图数据,并使用 GNN 进行地震的震源表征和震级估计,验证了基于 GNN 的震级估计方法的可行性和优势。

综上所述,为了实现一种快速又准确的多台输入的地震震级估计方法,本文提出一种基于图注意力网络(graph attention networks, GAT)的地震震级估计新方法(GAT_M),通过将首台 P 波触发后 3 s 的地震加速度时程数据构造为地震图数据集,并引入注意力机制以弥补传统图神经网络中节点邻居权重不变的限制,通过为邻居节点分配可变的边权重,自适应地提取节点特征,旨在更加精准地捕捉震级相关的关键信息。为了验证所提出方法的效果,对于相同的测试数据集,在首台 P 波触发后 3 s,本文将 GAT_M 模型的预测结果与传统震级估计方法中的 Pd 方法进行了对比分析。对比结果证明了 GAT_M 模型在准确性和可靠性方面具有显著优势。本研究期望为地震预警地震级的快速、精确估计提供一种新的方法论,并为相关领域的研究提供新的理论和实践路径。

1 方法

1.1 图与图神经网络

图是一种特殊的数据结构,由节点和连接这些节点的边组成。一个图 G 可以表示为一个二元组 $G=(V,E)$,其中 V 是节点的集合, E 是边的集合。图可以用于表示对象之间的关系或网络结构,通常分为无向图和有向图。无向图中,边没有方向;有向图中,边有方向。

图神经网络(GNN)是一种用于处理图结构数据的神经网络架构。GNN 的核心思想是通过节点之间的边来传播和更新节点的信息或状态,使其能够捕捉到图中节点的邻接关系和整体结构特征^[34]。在 GNN 中,节点的状态通过邻居节点的状态更新,通常通过聚合函数(如求和、平均)实现,进而通过神经网络层更新当前节点的状态。经过多次迭代,每个节点的状态包含了其邻域内的信息,使得 GNN 可以对图中的节点或整个图进行有效的表示和预测^[35]。

1.2 图注意力网络

图注意力网络(GAT)是处理图结构数据的一种先进方法^[36]。GAT 的核心思想在于引入注意力机制来动态地加权图中节点的邻居,从而有效地捕获节点间的依赖关系。

GAT 中的注意力机制通过单层前馈神经网络和非线性激活函数(如 LeakyReLU)计算节点对的特征并生成标量值,通过 softmax 函数标准化得到最终注意力权重。每个节点所获得的新的特征向量是其邻居节点特征向量的加权和,注意力机制决定权重,允许模型关注最相关的邻居节点信息。

1.3 GAT_M 模型

本文提出的 GAT_M 模型架构如图 1 所示。它包含 8 层图注意力卷积层,其中第 1 层为输入层,最后 2 层为输出层,中间 6 层为隐藏层。输入层的神经元为输入的节点特征数量,隐藏层的神经元数量为 512,输出层的神经元数量为 1。输入层和隐藏层中的图注意力卷积层的注意力头数为 4,而输出层的注意力头数为 6。在每个图注意力卷积层之间使用 ReLU6 激活函数进行激活,并在最后的图注意力卷积层输出接上一个全局平均池化作为图读出层,将节点特征转化为最终预测结果。GAT_M 模型的网络结构如表 1 所示。其中,GATconv 为图注意力卷积层,ReLU6 为激活函数,Global_mean_pool 为全局平均池化层, N 表示台站数量,最终输出即为震级大小。

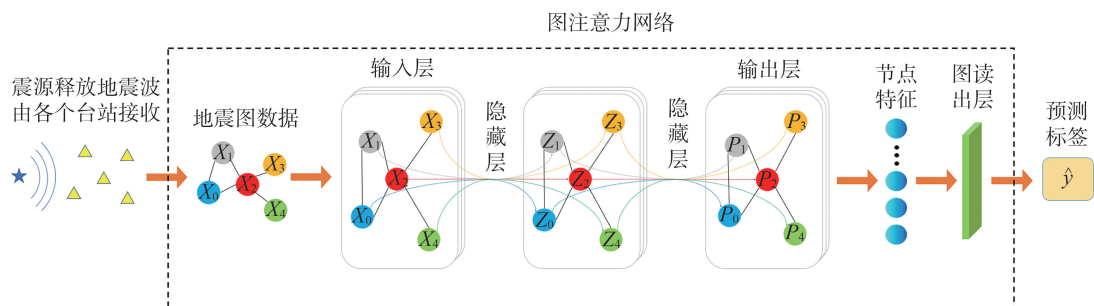


图 1 GAT_M 模型结构示意图

Fig. 1 Schematic diagram of the GAT_M model structure

表 1 GAT_M 模型的网络结构

Table 1 Network structure of the GAT_M model

| 网络层类型 | 网络层类别 | 注意力头数/个 | 神经元数量/个 | 输出特征 |
|-------|----------|---------|---------|--------------------|
| 输入层 | 地震图数据 | | | $N \times 30\,000$ |
| 特征提取器 | GATconv1 | 4 | 512 | $N \times 512$ |
| | ReLU6 | | | $N \times 512$ |
| | GATconv2 | 4 | 512 | $N \times 512$ |
| | ReLU6 | | | $N \times 512$ |
| | GATconv3 | 4 | 512 | $N \times 512$ |
| | ReLU6 | | | $N \times 512$ |

续表

| 网络层类型 | 网络层类别 | 注意力头数/个 | 神经元数量/个 | 输出特征 |
|-------|------------------|---------|---------|----------------|
| 特征提取器 | GATconv4 | 4 | 512 | $N \times 512$ |
| | ReLU6 | | | $N \times 512$ |
| | GATconv5 | 4 | 512 | $N \times 512$ |
| | ReLU6 | | | $N \times 512$ |
| | GATconv6 | 4 | 512 | $N \times 512$ |
| | ReLU6 | | | $N \times 512$ |
| | GATconv7 | 4 | 512 | $N \times 512$ |
| | ReLU6 | | | $N \times 512$ |
| 输出层 | GATconv8 | 6 | 1 | $N \times 1$ |
| 图读层 | Global_mean_pool | | | 1 |

2 数据

2.1 数据集及数据预处理介绍

本研究使用了日本防灾科学技术研究所 K-NET 强震观测台网在 2007 年 10 月至 2017 年 10 月收集到的地震数据^[37]。首先筛选出采样率为 100 Hz,震级在 3 级到 8 级之间,且震源深度小于 10 km 的地震数据。经过筛选后得到的地震数据一共有 2 909 次。图 2(a)展示了地震震中的地理分布。该图中,震中位置以圆点标识,圆点的颜色深浅变化直观地反映了震级的大小差异,从而提供了清晰的震级分布情况。图 2(b)则展示了地震监测网络的台站布局。在图 2 中,地震台站用三角形表示,主要城市以红色五角星图形明确标出,以便于快速识别和关联地震台站与周边城市之间的相对位置关系。这些地震数据由 1 004 个台站记录,总共包括 44 668 组三分量加速度信号。

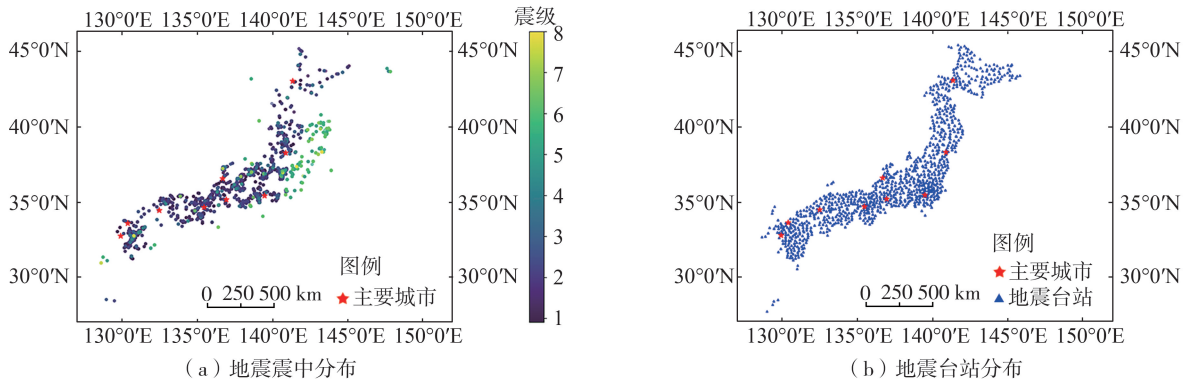


图 2 地震数据的震中和台站分布图

Fig. 2 Distribution of earthquake epicenters and stations for seismic data

本文的数据预处理步骤如下:

1) 采用马强等^[38]提出的长/短时平均方法与池赤准则相结合的方式对每个台站的 P 波到时进行自动拾取。在此基础上,再进行一次人工校对,将自动拾取失败的记录进行人工拾取 P 波到时,以保证可靠性。为了兼顾时效性和准确性,本研究使用 P 波触发后的 3 s 时间窗内的地震数据作为节点特征进行深入分析。

2) 对 3 s 时间窗内的加速度时程进行基线校正和消除仪器响应,并对数据进行归一化处理,表达式为

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

式中: X_{\min} 和 X_{\max} 分别为加速度时程中的最小值和最大值; X 为原始数据值; X_{norm} 为归一化后的数据值。此步骤将所有数据在不改变分布的情况下转换到同一范围内,可以有效提高模型训练效率和精度。

3) 将每个台站记录的三分量加速度时程转化为矢量模,其表达式为

$$W = \sqrt{W_{\text{EW}}^2 + W_{\text{NS}}^2 + W_{\text{UD}}^2} \quad (2)$$

式中: W_{EW} 、 W_{NS} 和 W_{UD} 分别为东西方向、南北方向以及垂直方向的加速度时程; W 为合成后得到的加速度时

程。虽然此合成过程可能导致部分信息的损失,但它在降低计算复杂性方面起到了显著作用。

图3展示了预处理后的加速度时程示例,其中包含来自5个台站的记录。横轴表示时间(以s为单位)。每个台站的记录分别编号为1~5。图中2条垂直虚线框选的部分表示P波触发后3s时间窗内的加速度时程。

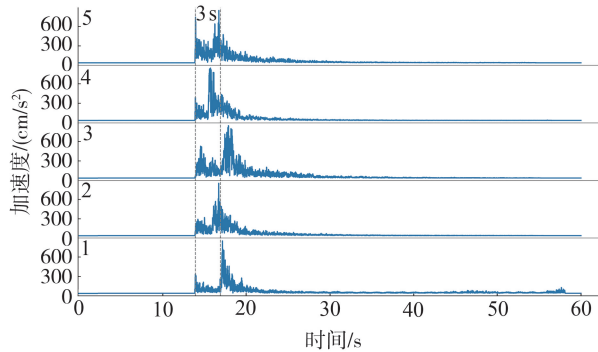


图3 预处理后的加速度时程示例图

Fig.3 Sample diagram of acceleration time history after preprocessing

2.2 地震图数据构造

本研究将地震数据用图 $G=(V,E,A)$ 进行表示。下面将分别对图中的3个属性 V,E,A 进行介绍。

1) V 表示节点的集合。将一个地震中的各个台站看作一个节点,台站收集到加速度时程数据看作节点特征,得到 $V=\{v_i\}, i=1, \dots, N$, 其中 N 为台站数量, v_i 为第 i 个台站收集到的加速度时程数据。特别地,图注意力网络的特征传递是通过邻居节点的特征聚合实现的,并不限制邻居节点的数量,因此,图注意力网络的输入不受台站数量的约束。本研究将一次地震中所有台站记录到的地震动记录用于构造一个地震图数据。

2) A 表示邻接矩阵。利用邻接矩阵来表示地震图数据集中各个台站的空间位置关系。假如一个地震由 N 个台站收集信息,则邻接矩阵的维度为 $N \times N$ 。邻接矩阵的元素只有0或1,其中矩阵元素 $a_{ij}=1$ 表示第 i 个台站与第 j 个台站的空间距离 d_{ij} 小于阈值 r ,代表二者在空间上相邻,需要综合考虑二者特征。反之,若矩阵元素 $a_{ij}=0$,则认为2个台站在空间上不相邻。邻接矩阵的判别公式为

$$a_{ij} = \begin{cases} 1, & \text{if } d_{ij} < r \\ 0, & \text{if } d_{ij} > r \end{cases} \quad (3)$$

本文中邻接矩阵的计算方法:

- ①读取加速度时程数据,得到台站的经纬度信息;
- ②通过将经纬度转换成弧度,使用 haversine 公式计算台站之间的空间距离:

$$H(\theta) = \sin^2\left(\frac{\theta}{2}\right) \quad (4)$$

$$d = 2r \cdot \arcsin(\sqrt{H(\Delta\phi) + \cos\phi_1 \cos\phi_2 H(\Delta\lambda)}) \quad (5)$$

式中: $H(\theta)$ 为 haversin 公式; d 为两点间的距离; r 为地球的平均半径,大约为 6371 km; ϕ_1, ϕ_2 为两点的纬度; $\Delta\phi$ 为两纬度之差; λ_1, λ_2 为两点的经度; $\Delta\lambda$ 为两经度之差。

3) 根据式(3)计算邻接矩阵。由地震预警系统的台站布设间隔通常为 15~20 km,本文选定 20 km 作为阈值,即第 i 个台站与第 j 个台站的空间距离 d_{ij} 小于 20 km,则 a_{ij} 记作 1。邻接矩阵对角线上的元素根据邻接矩阵的定义表示为该节点与自身的距离,均记作 1。

4) E 表示边的集合,即 $E=\{e_{ij} | a_{ij}=1\}$, 如果 $a_{ij}=1$, 则 $e_{ij} \in E$ 。其中, $e_{ij} \in E$ 意味着第 i 个节点与第 j 个节点间存在边关系,可以将第 i 个节点与第 j 个节点看作邻居节点。图4展示了一个拥有5个台站的地震数据转化为地震图数据的示意图。本研究使用首台P波触发后3s的加速度时程数据作为每个台站的节点特征。

鉴于各个地震记录所涵盖的台站数量并不统一,本文统计了不同台站数量的地震事件数目,具体数据见表2。为了增强预测模型的精准度,选取台站数量超过3个的地震事件作为地震图数据集进行模型的训练与测试。数据集中包含的震级与震中距的分布情况如图5所示。最终,本研究通过随机化处理,将地震图数据集分割为训练集与测试集两部分,并确保两者的比例控制在4:1。以上述方式准备的数据集,旨在实现对震级预测模型准确性的优化。

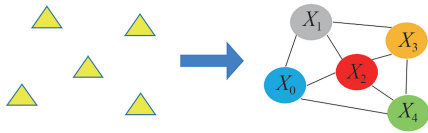


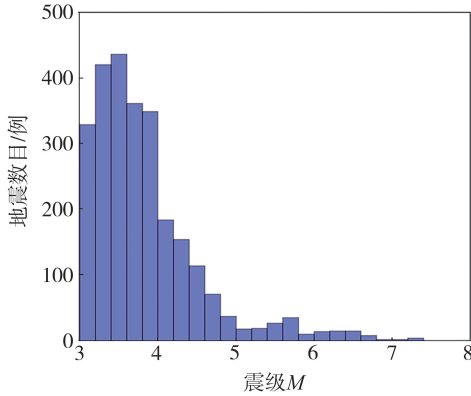
图 4 地震数据转化为地震图数据的示意图

Fig. 4 Schematic diagram of conversion of earthquake data into seismic map data

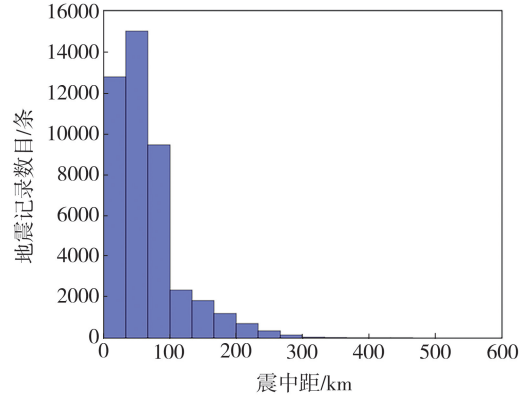
表 2 不同台站数量的地震数目分布

Table 2 Number distribution of earthquakes with different numbers of stations

| 台站数量 x | 地震数目/例 | 占比/% |
|----------------|--------|------|
| $x < 3$ | 285 | 10 |
| $3 \leq x < 5$ | 440 | 15 |
| $5 \leq x < 8$ | 543 | 19 |
| $x \geq 8$ | 1 641 | 56 |



(a) 震级分布



(b) 震中距分布

图 5 震级和震中距分布

Fig. 5 Distribution map of earthquake magnitude and epicentral distance

3 实验及结果分析

3.1 实验准备及参数确定

本文实验基于 Python 中的 Pytorch 库与 PyTorch Geometric 框架实现, CPU 为 Intel (R) Core (TM) i7-10700, GPU 为 NVIDIA GeForce GTX3060Ti, 运行内存为 16 GB。GAT_M 模型的超参数选取为: batch size = 16, epoch = 15, 损失函数采用平均绝对值误差函数 (mean absolute error, MAE), 优化器采用自适应矩估计算法 (adaptive moment estimation, Adam), 学习率为 5×10^{-4} 。同时 GAT_M 模型也使用了 L2 正则化技术, 在训练神经网络时限制模型权重大小从而减少过拟合的风险, L2 正则化的偏移量为 3×10^{-5} 。模型的超参数设置见表 3。

3.2 测试结果分析与评价

基于首台 P 波触发后 3 s 时间内的加速度时程数据, 将地震中所有台站记录到的地震动记录构造地震图数据并以 4:1 的比例划分训练集和测试集, 本文使用 70% 的训练集输入到 GAT_M 模型中进行模型训练, 同时使用 30% 的训练集作为验证训练效果的验证集, 模型训练和验证的损失变化曲线如图 6 所示。由图可知, 在经过 15 个 epoch 训练后, 训练和验证的损失均接近于 0, 不存在欠拟合的情况。

在网络得到良好的训练后, 本文使用测试集测试了 GAT_M 模型的性能。图 7 为测试集上真实震级与预测震级的比较以及误差的分布情况, 其中图 7(a) 为测试集上真实震级与 GAT_M 模型估算震级的比较图, 图 7(b) 为预测误差分布图。由图 7 可知, 真实震级和预测震级总体上是一致的, 并且大致遵循线性相

表 3 GAT_M 模型的超参数设置

Table 3 Hyperparameter setting of the GAT_M model

| 超参数类型 | 超参数值 |
|------------|--------------------|
| batch size | 16 |
| epoch | 15 |
| 损失函数 | MAE |
| 优化器 | Adam |
| 学习率 | 5×10^{-4} |
| L2 正则化 | 3×10^{-5} |

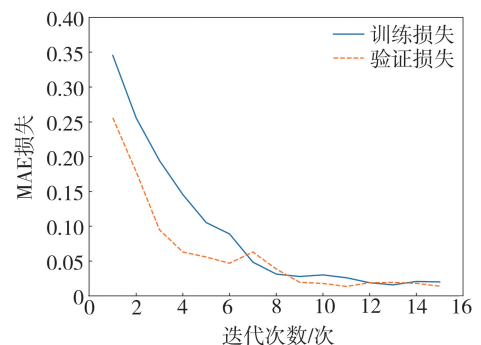


图 6 训练和验证过程中的损失曲线

Fig. 6 Loss curves during training and validation processes

关。通过计算得到预测震级的平均误差和标准差分别为 -0.077 和 0.40 震级单位,这说明 GAT_M 模型拥有较好的预测效果。 R^2 得分是一种易于计算和非常直观的用于度量相关性的指标,已经广泛用于回归模型中评估预测值与真实值的符合程度,其计算公式为

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \quad (6)$$

$$SS_{\text{res}} = \sum_i (y_i - \hat{y}_i)^2 \quad (7)$$

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2 \quad (8)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (9)$$

式中: y_i 为样本真实值; \hat{y}_i 为样本预测值; \bar{y} 为真实值的平均值; SS_{tot} 为总离差平方和,它表示样本总体偏离情况; SS_{res} 为残差平方和,它表示模型未能解释的偏离情况。由式(6)可得, R^2 得分一般在 $0 \sim 1$ 之间, R^2 得分越接近 1 说明模型拟合程度越高。通过计算得到 GAT_M 模型的 R^2 得分为 0.72 ,这表明模型的拟合程度很高。

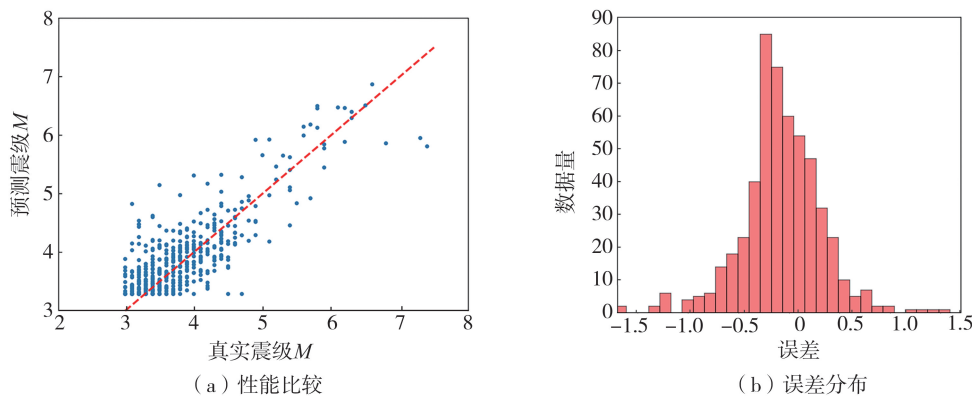


图7 GAT_M模型的测试结果表明

Fig. 7 Performance diagram of the test results of the GAT_M model

图8展示了 GAT_M 模型预测误差与震级之间的关系。通过按 0.5 个震级单位的间隔,取 ± 0.25 个震级单位的范围内的数据误差求平均,得到各震级段上的平均误差。分析图8可得,在小震级段($3 \sim 4$ 级的地震)和中震级段($4 \sim 6$ 级的地震)上,GAT_M模型表现出了较高的准确性。虽然在小震段上仍存在震级略微高估的现象。然而,当涉及到中震级段(>6 级的地震)时,预测效果出现了明显的下降,误差显著增大,导致预测震级与实际震级之间存在较大偏差,并且有明显震级低估的现象。性能下降的原因可能为大震样本数量的缺乏。由于大震发生的频率相对较低,因此在地震数据集中,大震的地震样本数量远少于小震和中震。样本数量不足导致模型难以准确学习这类地震的特征和规律,进而影响了预测准确度。

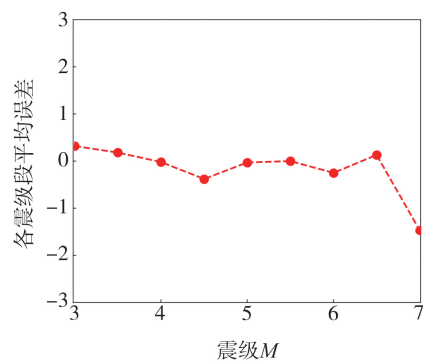


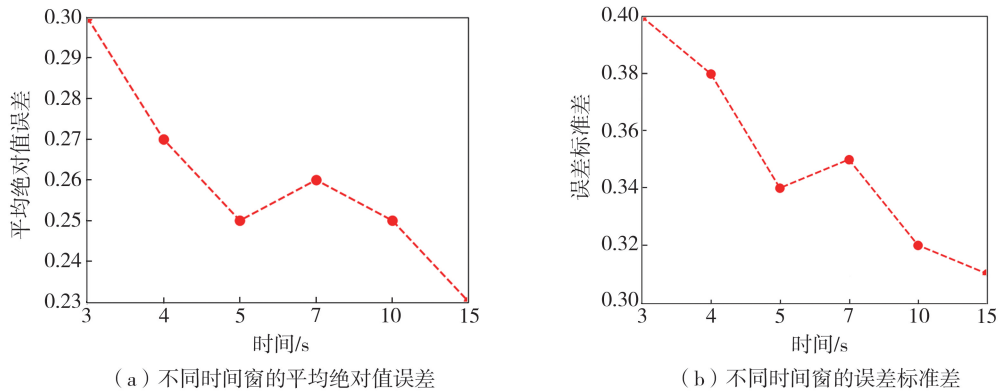
图8 GAT_M模型的预测结果误差与震级段的关系

Fig. 8 Relationship between the prediction error of the GAT_M model and the magnitude segment

此外,大震级地震的复杂性和变异性也可能是导致预测困难的一个因素。与小震和中震相比,大震的地质和地球物理过程更加复杂,这使得预测模型需要处理更多的不确定性和变量。因此,尽管 GAT_M 模型在小震和中震预测上表现良好,但在大震级段上仍需进一步优化和调整,以提高对这类地震的预测准确度。

在测试不同输入加速度时程时长对模型预测准确性的影响时,本文分别用测试集中的首台 P 波触发后 $3, 4, 5, 7, 10, 15$ s 的加速度时程数据作为节点特征构造地震图数据并进行测试,得到了如图9所示的结果。其中图9(a)为在首台触发后不同时间窗内预测结果的平均绝对误差的变化关系,而图9(b)为首台触发后不同时间窗内预测结果的误差标准差的变化趋势图。由图9可知,随着首台 P 波触发后时间窗的增加,平

均绝对误差和误差标准差总体呈下降趋势,尽管在某些时间点上有所轻微的上升。这表明随着时间窗的延长,预测的准确度在不断提高。



(a) 不同时间窗的平均绝对值误差

(b) 不同时间窗的误差标准差

图 9 预测误差与首台触发后的时间关系

Fig. 9 Relationship between prediction error and the time after the first trigger

本研究进一步探讨了台站数量对于 GAT_M 模型预测性能的影响。通过保持输入的加速度时程时长为 3 s 不变,仅使用台站数量超过特定阈值的震数据构造地震图数据,并以 4:1 的比例划分训练集和测试集。使用训练集来训练模型,并使用测试集分析了模型性能的表现,如图 10 所示。由图 10 可知,在使用台站数量大于 3 的样本时,GAT_M 模型的性能处于较低水平。然而,当使用台站数量大于 5 的样本时,模型性能显著提升,并且在使用台站数量大于 8 的样本时达到顶峰。这表明当台站数量增至一定数目时,对 GAT_M 模型预测效果的提升是显著的。然而,当台站数量继续增加时,模型性能呈现出下降趋势。这一现象可能表明,在达到最优台站数量之后,继续增加台站并未能引入额外的有价值信息,反而可能引入了噪声或导致模型过拟合,从而对模型的泛化能力产生负面影响。并且由于训练时放弃了台站数量较少的样本,使得模型的训练效果降低,导致模型整体性能下降。此发现对于确定震级估计模型中最佳的台站数量具有重要的指导意义,同时也提示了在震级估计模型设计中需平衡数据量与数据质量之间的关系。

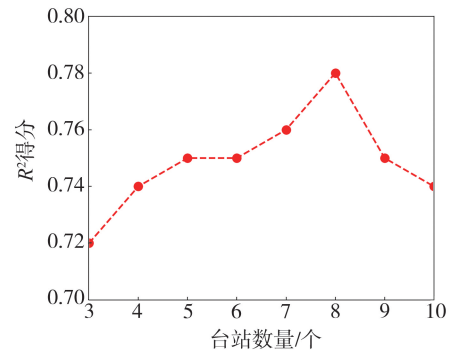


图 10 使用不同台站数量的样本训练出的模型性能表现图

Fig. 10 Performance of models trained with samples of different numbers of stations

3.3 与传统方法的对比分析

传统的地震预警系统中所使用的震级估计方法通常根据 P 波到达后的几秒内的地震动记录计算特征参数,然后建立特征参数与震级的经验关系,基于这个经验关系对震级进行快速估算^[7]。本文从传统震级估计方法中选取了峰值位移(Pd)方法与 GAT_M 模型进行对比分析。Pd 方法最早由 WU 和 ZHAO 提出^[39],目前被广泛运用于震级估计中^[7,22,38]。Pd 方法的计算公式为

$$\log(P_d) = A + BM + C\log(R) \quad (10)$$

式中: P_d 为峰值位移; M 为震级; R 为震源距; A 、 B 和 C 分别为拟合参数。

本文采用了与 GAT_M 模型相同的未经预处理的地震数据集,该数据集包括用于训练和测试的各台站的加速度时程数据。首先,对训练集和测试集的各个加速度时程进行 P 波到达时间拾取,并计算 P 波触发后 3 s 时间窗内的峰值位移 P_d 值。然后,从训练集中提取出 P_d 值,与震级 M 和震源距 R 一并代入式(10)中进行计算,得出最佳拟合参数: $A=1.299\pm 0.045$, $B=-0.0144\pm 0.0082$, $C=0.8120\pm 0.011$ 。在参数确定之后,使用测试集数据计算出的 P_d 值和震源距 R 代入式(10)来预测单个台站的震级。最后将每个地震事件中所有参与台站的预测震级进行汇总并计算其平均值,以此作为地震事件的最终预测震级。图 11 展示了 Pd 方法的预测性能。其中图 11(a)为相同数据集上真实震级与 Pd 方法估算震级的比较图,图 11(b)为 Pd 方法预测误差分布图。

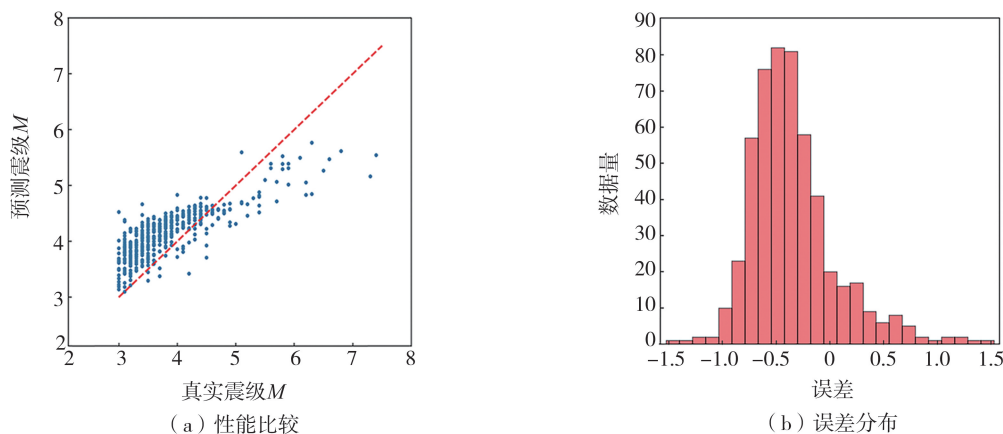


图 11 Pd 方法的测试结果表明

Fig. 11 Performance of the test results of the Pd method

由图 11 可知,在首台 P 波触发后 3 s, Pd 方法在预测小震(低震级地震)时往往估计过高,而在大震(高震级地震)的情况下则偏向低估。这会导致对震级的估计出现系统性偏差,从而影响预测结果的准确性和可靠性。相比之下,图 6 中的 GAT_M 模型的表现显著优于传统的 Pd 方法。GAT_M 模型的数据基本分布在 1:1 比例线周围,这表明该模型在不同震级范围内都具有一定的拟合效果,并且在预测小震和大震时效果均优于 Pd 方法。通过比较图 10(b) 和图 6(b) 可以发现, GAT_M 模型的误差基本集中在 ± 0.5 个震级单位以内,而 Pd 方法的误差则主要集中在 $[-1, 0]$ 之间,说明 GAT_M 模型在改善震级低估方面效果显著。基于相同的测试数据集,在首台 P 波触发后 3 s 计算得到 Pd 方法和 GAT_M 模型的各项性能指标如表 4 所示。对比可以发现, GAT_M 模型在 MAE、MSE 和 R^2 这 3 个平均指标上均优于 Pd 方法。综上所述,相比 Pd 方法, GAT_M 模型具有更高的震级预测准确性,这在及时响应和减少地震灾害风险中是至关重要的。

表 4 Pd 方法与 GAT_M 模型的性能比较

Table 4 Performance comparison between the Pd method and the GAT_M model

| 性能评价指标 | Pd 方法 | GAT_M 模型 |
|--------|-------|----------|
| MAE | 0.47 | 0.26 |
| MSE | 0.29 | 0.13 |
| R^2 | 0.39 | 0.72 |

4 结论及讨论

本文提出了一种基于图注意力网络的多台站地震预警震级估计新模型,利用日本 K-NET 强震台网记录的加速度时程数据构建了地震图数据集进行了模型训练和测试,得到如下结论:

1) GAT_M 模型在震级预测方面表现出高精度与稳定性,通过构造地震图数据输入模型进行训练和测试,得到 GAT_M 模型在测试集上的平均误差和标准差分别为 -0.077 和 0.40 ,且模型的决定系数 (R^2) 达到了 0.72 。GAT_M 模型在测试集上展示了良好的性能,反映出其优秀的泛化能力。

2) 本研究分析了震级、时间窗和台站数量对 GAT_M 模型性能的影响。GAT_M 模型对 3 级~6 级地震的预测具有较低的误差,而在预测 6 级以上的地震时误差增大。随着首台 P 波触发后时间窗的增加,预测误差总体呈下降趋势,表明预测精度在不断提高。此外,台站数量对模型性能有显著影响,实验发现选取台站数量大于 8 的地震数据训练模型时会得到最佳效果,实现数据量与数据质量的平衡。

3) 通过与传统 Pd 方法的预测结果进行对比分析,得到 GAT_M 模型的预测结果更加准确,并且在减少预测低估方面表现更佳。对 MAE、MSE 和 R^2 这三类性能指标的比较进一步证明了 GAT_M 模型的表现超越了 Pd 方法。

总体而言, GAT_M 模型的开发为地震预警领域带来了新的视角和工具。它不仅提高了震级估计的准确性,还为未来相关研究提供了一个有力的分析预测框架。对比 KUANG 等^[26]所提出的 MagNet 模型时,我们注意到其对输入台站数量设定了一个上限,即 20 个台站,且作为模型输入的加速度时程长度达到了 80 s。这种限制可能在实际地震监测和预警系统的部署中造成一定的局限性。本研究中提出的 GAT_M 模型在设计上则展现了显著的灵活性和适应性。GAT_M 模型不预设输入数据的台站数量上限,这样更贴近地震波传播的实际

动态。在地震波的初期传播阶段,通常仅有少数地震台站能够捕捉到数据;随着时间的推移,地震波进一步传播,能够收集数据的台站数量将持续增加,形成一个连续的数据流。GAT_M 模型能够在此基础上迅速构建初期的图数据,并能实时添加新的节点,这种输入特性与数据流的动态输入需求高度一致,使得地震预警系统能够充分利用所有可用的输入信息进行震级估计,从而提高预警的精度和可靠性。实时处理广泛的台站数据也意味着系统能够更全面地捕捉到地震波的传播特征,从而在预警时提供更精确的地震分析。因此,GAT_M 模型的这些设计特点不仅增强了地震预警系统的功能,也提升了其在实际应用中的效果和适应性。

尽管 GAT_M 模型已经展示了良好的效果,但仍有提升空间。首先,数据集中大震级地震样本的稀缺影响了模型的预测性能,在未来的研究中,我们会尝试数据增强技术或合成数据方法增加大震级样本数量,以平衡数据集中的震级分布,提高模型的性能。其次,本研究使用的数据集来自日本 K-net 台网,可能限制了其普适性和泛化能力,在未来的研究中我们会尝试使用其他地区的地震数据集,通过迁移学习的方式提高模型的泛化性,使得模型可以提高跨区域适用性。

致谢:感谢日本防灾科学技术研究所(NIED)为本文研究提供强震动数据。

参考文献:

- [1] 马强. 地震预警技术研究及应用[D]. 哈尔滨: 中国地震局工程力学研究所, 2008.
MA Qiang. Study and application on earthquake early warning [D]. Harbin: Institute of Engineering Mechanics, China Earthquake Administration, 2008. (in Chinese)
- [2] 宋晋东. 高速铁路运行控制用地震动参数及单台地震预警技术研究[D]. 哈尔滨: 中国地震局工程力学研究所, 2013.
SONG Jindong. Research on seismic ground motion indices for operation control and single station earthquake early warning applied for high-speed railway [D]. Harbin: Institute of Engineering Mechanics, China Earthquake Administration, 2013. (in Chinese)
- [3] ARANDA J M E, JIMENEZ A, IBARROLA G, et al. Mexico city seismic alert system[J]. *Seismological Research Letters*, 1995, 66(6): 42-53.
- [4] 司政亚, 庄建仓, 蒋长胜. 多个地震台网相互融合测定震级的贝叶斯算法[J]. *地球物理学报*, 2022, 65(6): 2167-2178.
SI Zhengya, ZHUANG Jiancang, JIANG Changsheng. A Bayesian algorithm for magnitude determination by merging multiple seismic networks[J]. *Chinese Journal of Geophysics*, 2022, 65(6): 2167-2178. (in Chinese)
- [5] HAVSKOV J, OTTEMOLLER L. Routine data processing in earthquake seismology[M]. Dordrecht: Springer Science & Business Media, 2010.
- [6] 彭朝勇, 杨建思, 薛兵, 等. 基于汶川主震及余震的预警参数与震级相关性研究[J]. *地球物理学报*, 2013, 56(10): 3404-3415.
PENG Chaoyong, YANG Jiansi, XUE Bing, et al. Research on correlation between early-warning parameters and magnitude for the Wenchuan Earthquake and its aftershocks[J]. *Chinese Journal of Geophysics*, 2013, 56(10): 3404-3415. (in Chinese)
- [7] 张红才, 金星, 李军, 等. 地震预警震级计算方法研究综述[J]. *地球物理学进展*, 2012, 27(2): 464-474.
ZHANG Hongcai, JIN Xing, LI Jun, et al. Review on magnitude estimation methods applied to earthquake early warning systems[J]. *Progress in Geophysics*, 2012, 27(2): 464-474. (in Chinese)
- [8] AKI K. Local site effects on weak and strong ground motion[J]. *Tectonophysics*, 1993, 218(1/2/3): 93-111.
- [9] YAMADA M, MORI J. Using τ_c to estimate magnitude for earthquake early warning and effects of near-field terms[J]. *Journal of Geophysical Research: Solid Earth*, 2009, 114(B5). DOI:10.1029/2008JB006080.
- [10] KANAMORI H. Real-time seismology and earthquake damage mitigation[J]. *Annual Review of Earth and Planetary Sciences*, 2005, 33: 195-214.
- [11] ALLEN R M, BROWN H, HELLEWEG M, et al. Real-time earthquake detection and hazard assessment by ElarmS across California[J]. *Geophysical Research Letters*, 2009, 36(5). DOI: 10.1029/2008GL036766.
- [12] 何斌, 朱景宝, 李山有, 等. 基于阈值预警的 2022 年 1 月 8 日青海门源 M_s 6.9 级地震潜在破坏区估计[J]. *地震工程与工程振动*, 2022, 42(4): 1-7.
HE Bin, ZHU Jingbao, LI Shanyou, et al. Estimation of potential damage zone of Qinghai Menyuan M_s 6.9 earthquake on January 8, 2022 based on threshold early warning[J]. *Earthquake Engineering and Engineering Dynamics*, 2022, 42(4): 1-7. (in Chinese)
- [13] 赵明, 陈石, 房立华, 等. 基于 U 形卷积神经网络的震相识别与到时拾取方法研究[J]. *地球物理学报*, 2019, 62(8): 3034-3042.
ZHAO Ming, CHEN Shi, FANG Lihua, et al. Earthquake phase arrival auto-picking based on U-shaped convolutional neural network[J]. *Chinese Journal of Geophysics*, 2019, 62(8): 3034-3042. (in Chinese)
- [14] ZHU J, LI Z F, FANG L H. USTC-Pickers: A unified set of seismic phase pickers transfer learned for China[J]. *Earthquake Science*, 2023, 36(2): 95-112.
- [15] ZHOU L Q, ZHAO C P, ZHANG M, et al. Machine-learning-based earthquake locations reveal the seismogenesis of the 2020 M_w 5.0 Qiaojia, Yunnan earthquake[J]. *Geophysical Journal International*, 2021, 228(3): 1637-1647.
- [16] KUANG W H, YUAN C C, ZOU Z H, et al. Autonomous earthquake location via deep reinforcement learning[J]. *Seismological Research Letters*, 2024, 95(1): 367-377.
- [17] 胡安冬, 张海明. 机器学习在地震紧急预警系统震级预估中的应用[J]. *地球物理学报*, 2020, 63(7): 2617-2626.

- HU Andong, ZHANG Haiming. Application of machine learning to magnitude estimation in earthquake emergency prediction system[J]. Chinese Journal of Geophysics, 2020, 63(7): 2617-2626. (in Chinese)
- [18] 林彬华, 金星, 康兰池, 等. 基于卷积神经网络的地震震级测定研究[J]. 地球物理学报, 2021, 64(10): 3600-3611.
LIN Binhua, JIN Xing, KANG Lanchi, et al. The research of earthquake magnitude determination based on convolutional neural networks[J]. Chinese Journal of Geophysics, 2021, 64(10): 3600-3611. (in Chinese)
- [19] 王自法, 廖吉安, 王延伟, 等. 基于深层卷积神经网络的震级快速估算方法[J]. 地球物理学报, 2023, 66(1): 272-288.
WANG Zifa, LIAO Jian, WANG Yanwei, et al. A fast magnitude estimation method based on deep convolutional neural networks[J]. Chinese Journal of Geophysics, 2023, 66(1): 272-288. (in Chinese)
- [20] 朱景宝, 宋晋东, 李山有. 基于支持向量机的地震预警震级快速估算研究[J]. 振动与冲击, 2021, 40(7): 126-134.
ZHU Jingbao, SONG Jindong, LI Shanyou. Rapid magnitude estimation for earthquake early warning based on SVM[J]. Journal of Vibration and Shock, 2021, 40(7): 126-134. (in Chinese)
- [21] LIU H Y, LI S Y, SONG J D. Discrimination between earthquake P waves and microtremors via a generative adversarial network[J]. Bulletin of the Seismological Society of America, 2022, 112(2): 669-679.
- [22] 宋晋东, 余聪, 李山有. 地震预警现地 PGV 连续预测的最小二乘支持向量机模型[J]. 地球物理学报, 2021, 64(2): 555-568.
SONG Jindong, YU Cong, LI Shanyou. Continuous prediction of onsite PGV for earthquake early warning based on least squares support vector machine[J]. Chinese Journal of Geophysics, 2021, 64(2): 555-568. (in Chinese)
- [23] 朱景宝, 宋晋东, 李山有. 基于深度卷积神经网络的 2021 年 5 月 21—22 日云南漾濞地震和青海玛多地震震级估算[J]. 地球物理学报, 2022, 65(2): 594-603.
ZHU Jingbao, SONG Jindong, LI Shanyou. Magnitude estimation of Yunnan Yangbi earthquake and Qinghai Madoi earthquake on May 21—22, 2021 based on deep convolutional neural network[J]. Chinese Journal of Geophysics, 2022, 65(2): 594-603. (in Chinese)
- [24] 曾奎原, 朱景宝, 宋晋东, 等. 基于深度卷积神经网络的 2022 年日本福岛 7.4 级地震震级估算[J]. 地震工程与工程振动, 2022, 42(5): 1-8.
ZENG Kuiyuan, ZHU Jingbao, SONG Jindong, et al. Magnitude estimation for the 2022 M7.4 Fukushima earthquake in Japan based on deep convolutional neural network[J]. Earthquake Engineering and Engineering Dynamics, 2022, 42(5): 1-8. (in Chinese)
- [25] MOUSAVI S M, BEROZA G C. A machine-learning approach for earthquake magnitude estimation[J]. Geophysical Research Letters, 2020, 47(1): e2019GL085976.
- [26] KUANG W H, YUAN C C, ZHANG J. Network-based earthquake magnitude determination via deep learning[J]. Seismological Research Letters, 2021, 92(4): 2245-2254.
- [27] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains[J]. IEEE Signal Processing Magazine, 2013, 30(3): 83-98.
- [28] QIU J Z, TANG J, MA H, et al. DeepInf: Social influence prediction with deep learning[C] // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London United Kingdom. New York: ACM, 2018: 2110-2119.
- [29] 何昊晨, 张丹红. 基于多维社交关系嵌入的深层图神经网络推荐方法[J]. 计算机应用, 2020, 40(10): 2795-2803.
HE Haochen, ZHANG Danhong. Recommendation method based on multidimensional social relationship embedded deep graph neural network[J]. Journal of Computer Applications, 2020, 40(10): 2795-2803. (in Chinese)
- [30] 王婷, 朱小飞, 唐顾. 基于知识增强的图卷积神经网络的文本分类[J]. 浙江大学学报(工学版), 2022, 56(2): 322-328.
WANG Ting, ZHUXiaofei, TANG Gu. Knowledge-enhanced graph convolutional neural networks for text classification[J]. Journal of Zhejiang University (Engineering Science), 2022, 56(2): 322-328. (in Chinese)
- [31] 杨世刚, 刘勇国. 融合语料库特征与图注意力网络的短文本分类方法[J]. 计算机应用, 2022, 42(5): 1324-1329.
YANGShigang, LIU Yongguo. Short text classification method by fusing corpus features and graph attention network[J]. Journal of Computer Applications, 2022, 42(5): 1324-1329. (in Chinese)
- [32] VAN DEN ENDE M P A, AMPUERO J P. Automated seismic source characterization using deep graph neural networks[J]. Geophysical Research Letters, 2020, 47(17): e2020GL088690.
- [33] MCBREARTY I W, BEROZA G C. Earthquake location and magnitude estimation with graph neural networks[C] // 2022 IEEE International Conference on Image Processing (ICIP). Bordeaux, France. Piscataway, NJ, USA: IEEE, 2022: 3858-3862.
- [34] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and deep locally connected networks on graphs, 2nd Int[C] // Conf. on Learning Representations, ICLR 2014—Conference Track Proceedings. 2014.
- [35] ZHOU J, CUI G Q, HU S D, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.
- [36] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. Stat, 2017, 1050(20): 10-48550.
- [37] AOI S, KUNUGI T, NAKAMURA H, et al. Deployment of new strong motion seismographs of K-NET and KiK-net[C] // Geotechnical, Geological, and Earthquake Engineering. Dordrecht: Springer Netherlands, 2010: 167-186.
- [38] 马强, 金星, 李山有, 等. 用于地震预警的 P 波震相到时自动拾取[J]. 地球物理学报, 2013, 56(7): 2313-2321.
MA Qiang, JIN Xing, LI Shanyou, et al. Automatic P-arrival detection for earthquake early warning[J]. Chinese Journal of Geophysics, 2013, 56(7): 2313-2321. (in Chinese)
- [39] WU Y M, ZHAO L. Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning[J]. Geophysical Research Letters, 2006, 33(16): 271-284.