

任涛, 林梦楠, 陈宏峰等. 2019. 基于 Bagging 集成学习算法的地震事件性质识别分类. 地球物理学报, 62(1):383-392, doi: 10.6038/cjg2019M0380.

Ren T, Lin M N, Chen H F, et al. 2019. Seismic event classification based on bagging ensemble learning algorithm. *Chinese J. Geophys.* (in Chinese), 62(1):383-392, doi:10.6038/cjg2019M0380.

基于 Bagging 集成学习算法的地震事件性质识别分类

任涛¹, 林梦楠¹, 陈宏峰^{2*}, 王冉冉¹, 李松威¹, 刘晓雨², 刘杰²

¹ 东北大学软件学院, 沈阳 110819

² 中国地震台网中心, 北京 100029

摘要 地震台网在监测地震的同时记录到的非天然震动事件会对后续的科研和预报工作造成较大的影响, 因此快速准确的对天然震动事件与非天然震动事件加以区分就显得尤为重要. 本文针对传统人工方法识别地震事件性质的不足之处, 采用 Bagging 机器学习算法对地震事件性质进行区分. 首先选取震中距范围在 80~200 km 内的地震数据, 之后采用 AIC 算法自动识别 P 波到时, 进而用处理后的数据训练模型, 最后使用测试数据对模型进行评估, 准确率可达 85% 以上. 因此, 本文提出的方法可以有效地对天然震动事件与非天然震动事件加以区分.

关键词 地震事件分类; 频谱比值; 自相关系数; Bagging 算法

doi:10.6038/cjg2019M0380

中图分类号 P315

收稿日期 2018-07-04, 2018-11-22 收修定稿

Seismic event classification based on bagging ensemble learning algorithm

REN Tao¹, LIN MengNan¹, CHEN HongFeng^{2*}, WANG RanRan¹,

LI SongWei¹, LIU XiaoYu², LIU Jie²

¹ Software College of Northeastern University, Shenyang 110819, China

² China Earthquake Networks Center, Beijing 100029, China

Abstract The non-natural vibration events recorded by the Seismic Network while monitoring the earthquake will have a greater impact on the subsequent research and forecasting work. Therefore, it is particularly important to distinguish between natural earthquakes and non-natural vibration events quickly and accurately. In this paper, the Bagging machine learning algorithm is used to distinguish the nature of earthquake events, to improve the inadequacies of traditional artificial methods to identify the nature of earthquake events. Firstly, the seismic data with the epicenter distance in the range of 80~200 km is selected. Then, the AIC algorithm is utilized to automatically identify the arrival time of the P wave. After that, the processed data is used to train the model. Finally, the model is evaluated using the test data, and the accuracy rate is up to 85%. The method proposed in this paper can effectively distinguish between natural earthquakes and non-natural vibration events.

Keywords Seismic event classification; Spectrum ratio; Autocorrelation coefficient; Bagging algorithm

基金项目 国家自然科学基金资助项目(61473073, 61104074), 中央高校基本科研业务费(N161702001), 辽宁省高校优秀人才基金(LJQ2014028)资助.

第一作者简介 任涛, 男, 博士, 教授, 主要从事人工智能相关研究. E-mail: chinarentao@163.com

*** 通讯作者** 陈宏峰, 男, 高级工程师, 长期从事地震监测业务研究. E-mail: chf@seis.ac.cn

0 引言

地震是一种由板块运动引起的自然现象,在地震监测的过程中,我们记录到了许多天然地震事件,同时也记录到了许多由于人类活动而引起的地表震动事件,如:爆炸、塌陷、滑坡、飞行物坠落等.二者的震源性质有所不同,天然地震大多发生于地下几公里至几十公里的地壳深处,其震源是非对称剪切源(曾融生等,2000),而非天然震动事件种类多样,其中,爆破事件为膨胀源,滑坡事件主要为单力偶源,塌陷事件随着不同塌陷方式又有区别等.

随着地震台网中台站密度的增加,人类对震动事件的监测能力也逐步增强,记录到的小规模震动事件的数量也呈指数级上升(龙锋和阮祥,2017).在这些震动事件中,天然地震与非天然震动事件相互参杂,且两者波形较为相似,需要一定的经验才能将二者区分开来.一直以来,地震台网工作人员都是依靠经验对天然震动事件与非天然震动事件加以区分,受主观因素影响较多,难以实现统一判别标准.识别过程中要求工作人员掌握较为丰富的实践经验和较为扎实的地震学功底,且事件性质判断所需时间较长,效率和准确率不高.由于部分工作人员在事件性质分析识别方面经验相对不足,造成波形性质判断不准确的现象时有发生,这些误判事件如果不能及时剔除,会混淆我们的地震目录,对后续的科研和预报工作造成较大的影响.因此提出一种快速、准确对天然地震与非天然震动事件进行区分的方法就显得尤为重要.

在地震事件性质分类过程中,如何提取出有效特征是分类的关键.自20世纪50年代开始,国内外在地震事件性质识别方面进行了广泛和深入的研究,并提出了多种识别判据,主要有,P波初动(康英等,2001)、震源深度、体波震级与面波震级之比、P波初动振幅与P波最大振幅比、勒夫波和瑞利波振幅比、P波与S波谱振幅比、P波与勒夫波谱振幅比、倒谱、小波变换(刘希强等,2003;边银菊,2002;杨选辉等,2005;和雪松等,2006;Taylor,2011)、瞬态谱等(Giudicepietro et al.,2017),但是由于地震信号本身的复杂性,有些识别判据的识别效果或适用性仍有不足,造成某些判据仅仅适用于一些特定的地区和台站记录.

机器学习理论和应用研究始于1986年(Perol et al.,2018),其在图像识别(Bai et al.,2018)、信号

处理、预测评估、组合优化及知识工程等领域有着广泛的应用.近年来机器学习方法开始被用来对地震波形数据进行分析 and 处理. Dowla 等利用多层感知器(MLP)神经网络识别天然地震与地下爆炸(Giudicepietro et al.,2017);黄汉明(黄汉明等,2010;王婷婷和边银菊,2011;毕明霞等,2011)等使用提取的特征值作为输入数据,使用支持向量机的方法对天然地震与爆破进行识别.

随着人工智能的突飞猛进,如何把日新月异的计算机前沿技术应用于地震事件分类中,从而提高识别的准确率、稳定性仍是当前需要研究的课题之一.本文首先从波形记录中提取出有效、适用的特征值,在此基础上用决策树模型作为基分类器,采用Bagging集成方法将其集成为一个强分类器,从而提高单个分类器的识别准确率.

1 事件性质分类算法

根据样本策略,集成学习方法主要分为 Bagging、Boost(Pham et al.,2016; Breiman,1996; Freund and Schapire,1996)两种. Bagging 算法是一种用来提高弱分类器准确率的集成方法,这种方法基本思想是对训练集实施有放回的随机样本重新抽样,形成多个与训练集规模相近但各不相同的训练子集,在此基础上形成多个基分类器,在多个基分类器的基础上使用多数投票原则(Bayar and Stamm,2018).它利用训练集中随机抽取出的子集来训练每个模型,这有助于降低方差并有助于避免过拟合现象的发生.其算法过程如下:

(1) 从原始样本集中抽取训练集. 每轮从原始样本集中使用 Bootstrapping 的方法抽取 n 个训练样本(在训练集中,有些样本可能被多次抽取到,而有些样本可能一次都没有被抽中). 共进行 k 轮抽取,得到 k 个训练集(k 个训练集相互独立).

(2) 每次使用一个训练集得到一个模型, k 个训练集共得到 k 个模型(注:根据具体问题采用不同的分类或回归方法,如决策树、神经网络等).

(3) 对分类问题:将上一步得到的 k 个模型采用投票的方式得到分类结果. 对回归问题,计算上述模型的均值作为最后的结果.

为了让大家更好地理解 Bagging 方法,这里提供一个例子. X 表示一维属性, Y 表示类标号(1 或 -1)测试条件,当 $x \leq k$ 时, $y = ?$, 当 $x > k$ 时, $y = ?$, k 为最佳分裂点,属性 x 对应的唯一正确的 y

类别如表 1 所示.

表 1 属性 x 对应的 y 类别

Table 1 Attribute corresponding category										
X	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Y	1	1	1	-1	-1	-1	-1	-1	1	1

现在进行 5 轮随机抽样,抽取结果如表 2 所示,其中**黑体数字**表示未命中.

表 2 随机抽取结果
Table 2 Randomly drawn results

第一轮: $k=0.75; x \leq k, y=-1; x > k, y=1$; 准确率=70%										
X	0.1	0.4	0.5	0.6	0.6	0.7	0.8	0.8	0.9	0.9
Y	1	-1	-1	-1	-1	-1	-1	-1	1	1
第二轮: $k=0.65; x \leq k, y=-1; x > k, y=1$; 准确率=60%										
X	0.1	0.2	0.3	0.4	0.5	0.8	0.9	1	1	1
Y	1	1	1	-1	-1	-1	1	1	1	1
第三轮: $k=0.35; x \leq k, y=1; x > k, y=-1$; 准确率=90%										
X	0.1	0.2	0.3	0.4	0.4	0.5	0.7	0.7	0.8	0.9
Y	1	1	1	-1	-1	-1	-1	-1	-1	1
第四轮: $k=1; x \leq k, y=1; x > k, y=-1$; 准确率=50%										
X	0.1	0.1	0.2	0.5	0.6	0.7	0.7	0.8	0.9	0.9
Y	1	1	1	-1	-1	-1	-1	-1	1	1
第五轮: $k=0.4; x \leq k, y=1; x > k, y=-1$; 准确率=70%										
X	0.1	0.1	0.2	0.5	0.6	0.6	0.6	1	1	1
Y	1	1	1	-1	-1	-1	-1	1	1	1

每一轮随机抽样后,都会生成一个分类器,然后将五轮分类融合,融合结果如表 3 所示.

表 3 分类融合结果
Table 3 Classification fusion results

轮	k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	0.75	1	-1	-1	-1	-1	-1	-1	-1	1	1
2	0.65	1	1	1	-1	-1	-1	1	1	1	1
3	0.35	1	1	1	-1	-1	-1	-1	-1	-1	1
4	1	1	1	1	-1	-1	-1	-1	-1	1	1
5	0.4	1	1	1	-1	-1	-1	-1	1	1	1
和	—	5	3	3	-5	-5	-5	-3	1	3	5
符号	—	1	1	1	-1	-1	-1	-1	1	1	1
实际类	—	1	1	1	-1	-1	-1	-1	-1	1	1

对比符号和实际类,我们可以发现,在该例子中, Bagging 使得分类准确率高达 90%. 由此,总结 Bagging 方法的特点如下:

(1) Bagging 通过降低基分类器的方差,改善了

泛化误差.

(2) 其性能依赖于基分类器的稳定性,如果基分类器不稳定, Bagging 有助于降低训练数据的随机波动导致的误差. 如果稳定,则集成分类器的误差主要由基分类器的偏倚引起.

(3) 由于每个样本被选中的概率相同,因此 Bagging 并不侧重于训练数据集中的任何特定实例.

2 数据预处理

2.1 数据集的选取

本文从中国地震台网中心获取震动事件波形数据,用于天然地震和非天然地震的识别. 为了最大限度的提取出适用于不同台站记录的波形特征,本文从 2012—2015 年发生的 $M_L \geq 2.0$ 级震动事件中随机抽取 1058 次震动事件(含 4887 条台站纪录)作为研究对象,其中训练集包含天然地震事件 358 次(含 2394 条台站波形),其余为非天然地震事件. 测试集包含天然地震事件 272 次(含 785 条台站波形),其余为非天然地震事件(具体震动事件分布如图 1 所示).

2.2 选取数据质量较好的波形记录

地震台站在记录波形数据的过程中会受到各种外界因素的影响形成噪声,同时由于通讯问题或设备故障等因素会造成波形信号中断,这些都会给后期的信号处理带来影响. 因此,需要对采集到的原始地震波形进行预处理,去除断记、信噪比低的地震波形数据.

2.3 台站选择

本文选取震中距在 80~200 km 范围内的所有台站波形数据作为训练集和测试集. 在系统验证时选取震中距 80~200 km 范围内的所有台站波形数据进行识别.

2.4 P 波到时的拾取

在对地震事件性质进行分析识别时,需要首先拾取到 P 波到时. 拾取 P 波到时的方法有很多,如 STA/LTA 算法、瞬时频率算法、AIC 算法、动态特征算法(刘晗和张建中, 2014; Allen, 1978, 1982; Freund and Schapire, 1996; Baer and Kradolfer, 1987). 本文采用 AIC 算法进行 P 波到时拾取,并以拾取到的 P 波到时作为地震事件的起点(Maeda, 1985). 从拾取到的 P 波到时开始,截取之后 169 s 的事件波形作为训练和测试数据.

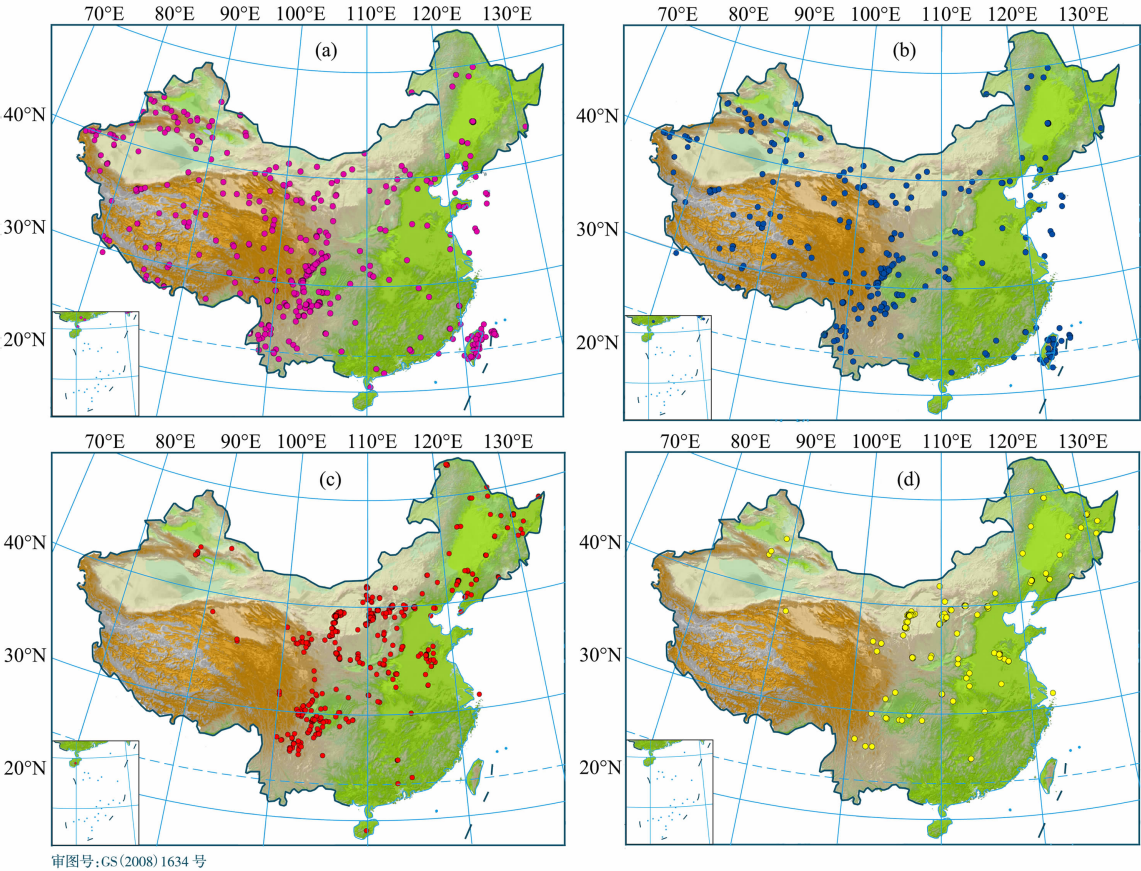


图 1 震动事件分布图
(a) 天然地震训练集事件分布图; (b) 天然地震测试集事件分布图; (c) 非天然震动训练集事件分布图;
(d) 非天然震动测试集事件分布图.

Fig. 1 Seismic data distribution map
(a) Natural seismic training set data distribution map; (b) Natural seismic test set data distribution map;
(c) Non-natural data training set data distribution map; (d) Non-natural data test set data distribution map.

3 波形特征提取

为了鉴别天然震动事件和非天然震动事件的波形差异和事件类型,本文提取天然地震和非天然地震波形的波形复杂度、谱比值、自相关系数、波形复杂度和自相关系数的综合比值作为波形特征来完成天然地震和非天然震动事件的区识别别.

3.1 波形复杂度(C)

波形复杂度特征最早是由英国的一个隶属于原子武器组织的研究小组提出的. 相关研究人员注意到,地下核爆炸所产生的 P 波波形相对简单——大振幅的 P 波持续一两个周期,紧随其后的是小振幅的尾波. 与此相应的天然地震的波形通常比较复杂,有很多相似振幅的波至,持续 35 s 或更长时间. 如果非天然震动事件源于复杂非均匀的环境中或是地震伴有大的应力降或者很快的破裂速度时,该判别

方法较为显著(韩绍卿等,2011). 针对二者 P 波能量衰减速度不同的特点,本文选择波形复杂度作为判断事件类型的特征值之一,其定义如下:

$$C = \left(\int_0^5 y^2(t) dt / 5 \right) / \left(\int_5^x y^2(t) dt / (x - 5) \right), \quad (1)$$

其中, $y(t)$ 为时域波形中 t 时刻的幅值,积分限中的 0 代表 P 波到达该台站的时刻,5 为经验系数,代表使用从 P 波初至开始 5 s 的波形段(经验表明,非天然震动事件 P 波能量衰减主要在 5 s 左右完成,而天然地震事件 P 波能量衰减周期相对较长), x 为 S 波与 P 波的到时差,本文选择中国地区 P 波和 S 波传播的平均速度为 $6 \text{ km} \cdot \text{s}^{-1}$ 和 $3.5 \text{ km} \cdot \text{s}^{-1}$ 进行计算,则 $x = \Delta / 3.5 - \Delta / 6$ (Δ 为震中距). 本文中选取震中距在 80 至 200 km 之间的台站进行计算.

波形复杂度度量的是 P 波初始段与 P 波尾波段能量的比值. 计算后得到不同波形复杂度取值范围所占百分比分布如图 2 所示.

由图 2 可知天然地震和非天然震动的波形复杂度取值在 $[0,5]$ 范围内,当波形复杂度取值在 0 至 1.4 范围内时,本次事件为天然地震事件的概率较大,当波形复杂度取值大于 1.4 时,本次事件为非天然震动事件的概率较大。

3.2 频谱比(SR)

经验表明,天然地震和非天然震动事件的频谱可能存在差异。图 3 分别代表了 2018 年 6 月 17 日甘肃酒泉市阿克塞县 M4.5 级地震时甘肃石包城台站(SBC)记录到地震波形的频谱分析和 2018 年 2 月 6 日 11 时 34 分 23 秒宁夏平罗地区爆破时内蒙巴彦浩特台站(BYT)记录波形的频谱分析。

本文采用波形数据频谱的比值作为特征值之一,其定义如下:

$$SR = \int_{L_1}^{H_1} |x(f)| \, df / \int_{L_2}^{H_2} |x(f)| \, df, \quad (2)$$

其中, $x(f)$ 为傅里叶频谱中频率为 f 处幅值的模; H_1 、 L_1 为波形频谱中频率较低部分的频率上限值和下限值,本文中取值为 $L_1 = 0, H_1 = 1, L_2 = 1, H_2 = +\infty$ 。计算后得到不同频谱比取值所占百分比情况如图 4 所示。

由图 4 可知天然地震和非天然震动的频谱比取值在 $[0,10]$ 范围内,由图 4b 可知当频谱比的取值在 0 至 0.97 范围内时,本次事件为非天然震动的概率较大,当频谱比的取值大于 0.97 时,为天然地震的概率较大。

3.3 自相关系数(R)

自相关系数度量的是同一事件在两个不同时期

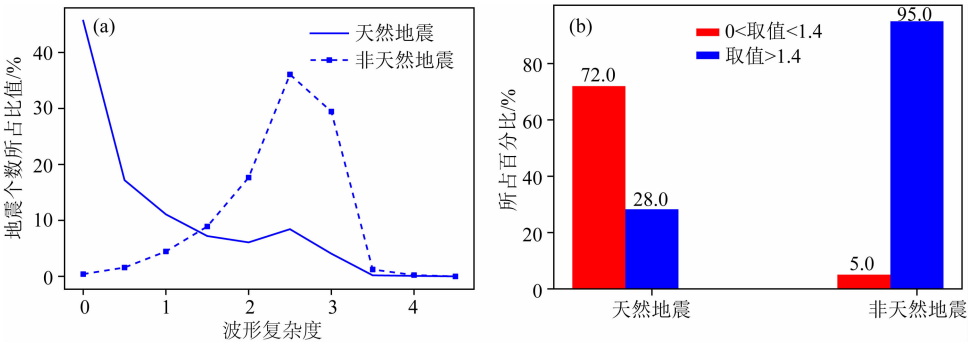


图 2 波形复杂度的分布特征
(a) 不同波形复杂度所占百分比; (b) 波形复杂度与事件类型的关系。
Fig. 2 Distribution characteristics of waveform complexity

(a) Percentage of different waveform complexity; (b) Relationship between waveform complexity and event type.

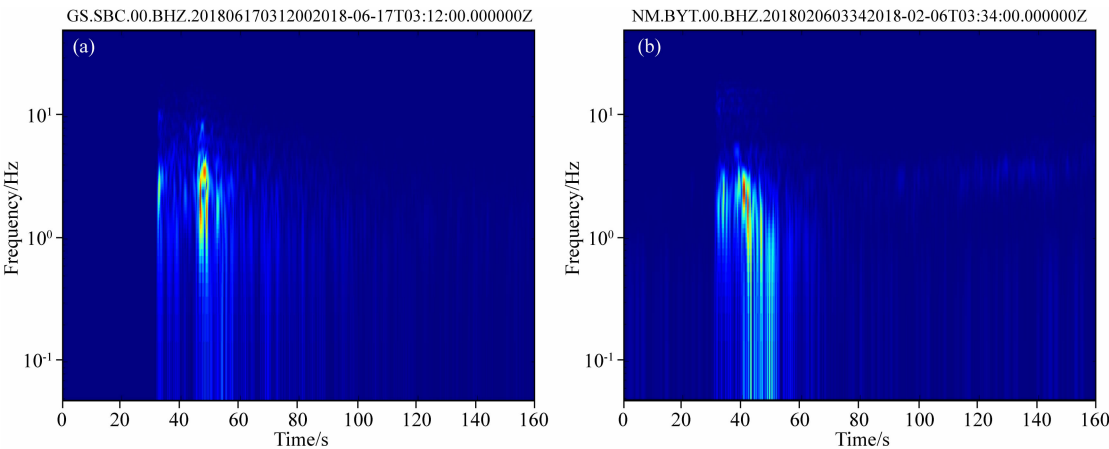


图 3 频谱分析图
(a) 2018 年 6 月 17 日甘肃酒泉市阿克塞县 M4.5 地震时甘肃石包城台站(SBC)记录到地震波形的频谱分析图;
(b) 2018 年 2 月 6 日 11 时 34 分 23 秒宁夏平罗地区爆破时内蒙巴彦浩特台站(BYT)记录波形的频谱分析图。
Fig. 3 Spectrum analysis diagram

(a) Earthquake waveforms recorded at the Shibaocheng station (SBC) in Gansu province at the M4.5 earthquake in Aksai County, Jiuquan, Gansu, June 17, 2018; (b) Spectral analysis of recorded waveforms at Bayanhot Station (BYT), Inner Mongolia, at 11:34:23 on February 6, 2018 at the blasting in the Pingluo of Ningxia Province.

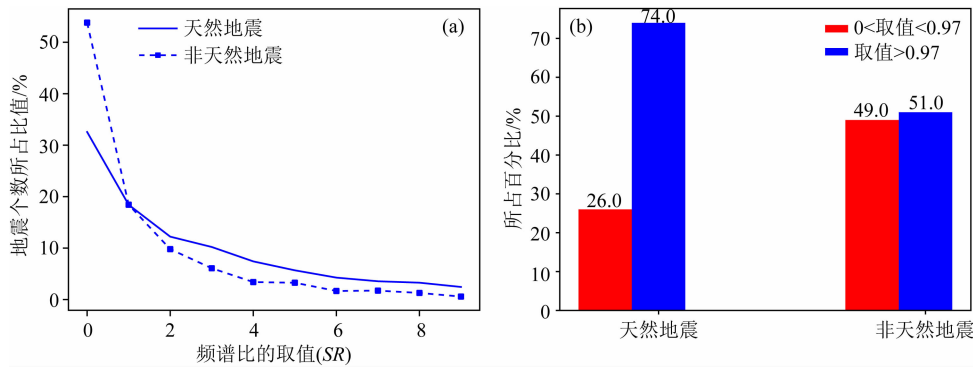


图 4 频谱比的分布特征
(a) 不同频谱比取值所占百分比; (b) 频谱比与事件类型的关系。

Fig. 4 Distribution characteristics of spectral ratio

(a) Percentage of different spectrum ratios; (b) Relationship between spectrum ratios and event types.

之间的相关程度,形象地讲就是度量自己过去的行为对自己现在的影响.由于大多数非天然震动都是能量瞬间爆发,幅度变化大,而天然地震相对于非天然震动衰减时间慢,幅度变化相对较小.这种幅度的变化可以反应在自相关系数的变化上.自相关系数的计算如下:

$$R = \frac{\sum_{i=1}^{n-h} (X_i - \eta)(X_{i+h} - \eta)}{\sum_{i=1}^n (X_i - \eta)^2}, \quad (3)$$

其中, n 为采样点个数, h 为计算自相关系数时波形时间差, η 为全部波形数据的均值.鉴于非天然震动事件震级一般不会大于 4 级,且 P 波衰减时间一般在 5 s 左右,所以本文选择用当前波形数据与 5 s 前的波形数据做自相关,希望通过 P 波衰减的速度反映出二者的差异,从而实现事件性质判断的目标.使用三分向地震数据,分别求出三个分向的自相关系数,然后将三分向的自相关系数计算结果求和,作为参与判定的特征值之一.计算后得到的自相关系数结果的取值及其所占百分比如图 5 所示.

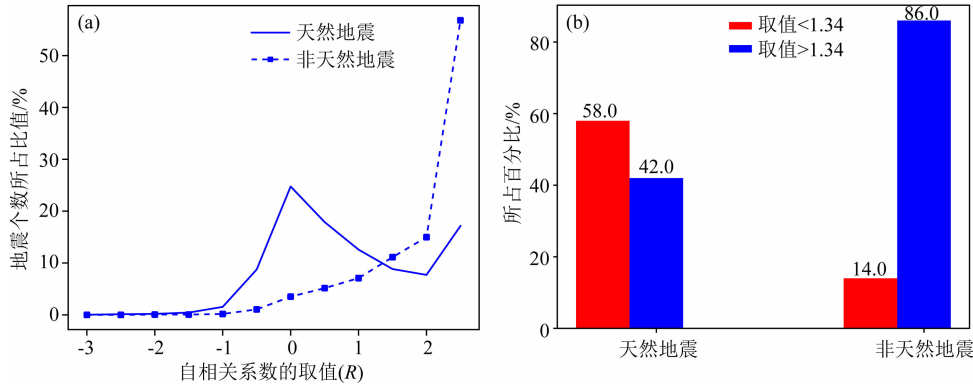


图 5 自相关系数的分布特征
(a) 不同自相关系数所占百分比; (b) 自相关系数与事件类型的关系。

Fig. 5 Distribution characteristics of autocorrelation coefficients

(a) Percentage of different autocorrelation coefficients; (b) Relationship between autocorrelation coefficients and event types.

由图 5 可知,天然地震和非天然震动的自相关系数取值主要集中在 $[-2, 3]$ 范围内.当自相关系数取值在 -2 至 1.34 范围内时,本次事件为天然地震的概率较大,当自相关系数取值在 1.34 至 3 范围内时,为非天然震动事件的概率较大,且非天然地震事件的自相关系数取值主要集中在 1.34 至 3 区间内.

3.4 波形复杂度和自相关系数的比值(Z)

波形复杂度、频谱和自相关系数分别代表了地震波形事件的某种特性,为了提高模型识别准确率,综合考虑波形复杂度和自相关系数,对同一事件的波形复杂度和自相关系数求比值,即综合比=波形复杂度/自相关系数,从而得到综合比计算公式如下:

$$Z = C/R. \quad (4)$$

计算后综合比值的分布和不同综合比取值所占百分比分布如图 6 所示.

由图 6 可知,综合比的取值在 $[-1, 3]$ 区间内.当综合比的取值在 -1 至 0.64 区间范围内时,本次事件为天然震动的概率较大;当综合比的取值大于

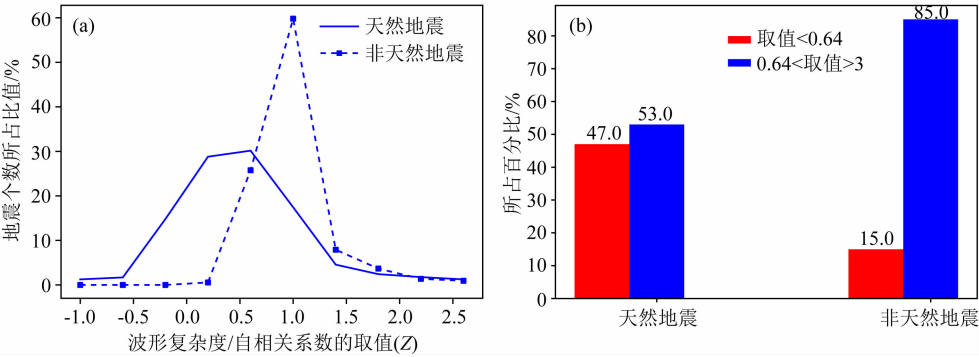


图 6 综合比的分布特征
(a) 不同综合比所占百分比; (b) 综合比取值与事件类型的关系。

Fig. 6 Distribution characteristics of composite ratios
(a) Percentage of different composite ratios; (b) Relationship between composite ratios and event types.

0.64 时,为天然地震事件的概率较大.且非天然震动事件的综合比取值主要分布在 0.64 至 3 区间.

4 事件性质判断

对于每次震动事件,首先计算出所选择震中距范围内所有符合规则台站的相关特征值,再送入模型中进行分类识别.结果判断时,采用少数服从多数的原则,以多数台站的分类预测结果为本次震动事件的最终类型判断结果.当预测结果为 1 和 -1 的台站个数相同时,以震中距最近台站的投票判定结果为准.

5 实验及分析

在 Windows 操作系统下,采用 Python 3.6 作为开发环境.分类模型用来识别区分天然地震事件还是非天然震动事件,也就是将模型的输出设为两类,其中输出为 1 代表是天然地震事件,输出为 -1 代表是非天然震动事件.由上述分析,本文基于决策树集成的地震事件识别流程如图 7 所示.

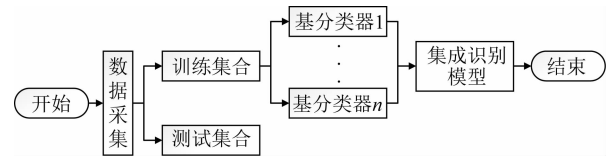


图 7 地震事件识别流程图

Fig. 7 Earthquake event identification flow chart

图 7 是基于本文设计实现的一种地震事件识别过程,数据处理模块对数据进行预处理及特征提取,形成输入向量.在此基础上通过 Bootstrap 采样得到训练子集并形成基分类器,得到集成识别模型,在

测试时采用多数投票方法对数据进行预测并输出.

基分类器的个数是对 Bagging 集成模型影响最大的参数,理论上,数据量越大,集成的基分类器个数应适当增加(Melki et al., 2018). 本文设计了 5 个对比实验,分别选取 10、60、100、150、200 个基分类器进行训练.并分别使用准确率(Accuracy)和召回率(Recall)两个度量值,用来评价结果的质量,一般来说:准确率=提取出的正确信息条数/提取出的信息条数,召回率=提取出的正确信息条数/样本中的信息条数.测试结果如图 8 所示.

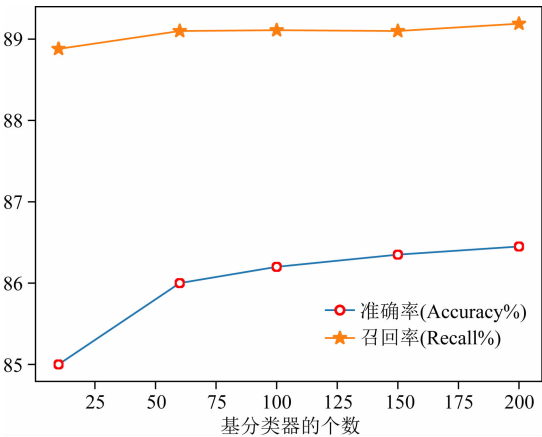


图 8 测试结果对比图

Fig. 8 Comparison of test results

由图 8 可知,不同基分类器个数的模型准确率均大于 85%,说明 Bagging 集成学习算法可以很好地对地震事件性质进行分类.同时,随着基分类器的不断增加,准确率和召回率均呈现不断上升的趋势,且在基分类器的个数为 200 个时,准确率和召回率均达到最高,为 86.62%. 因此本文选择 200 个决策树为基分类器的 Bagging 集成模型.

为验证本文集成学习模型的识别能力,利用

200 个基分类器的集成学习模型与支持向量机 (SVM) 和决策树 (Decision Tree) 进行非集成学习模型实验对比. 实验数据如表 4 所示.

表 4 模型准确率对比

Table 4 Comparison of model accuracy

实验方案	准确率 (Accuracy) (%)	召回率 (Recall) (%)
Ensemble_200	0.8662	0.8819
SVM	0.8428	0.8782
Decision Tree	0.8345	0.8630

表 4 中数据表明,这两种方法的准确率和召回率均小于本文所用模型,这是由于集成学习的主要思想是对新的实例进行分类的时候,把若干个单个分类器集成起来,通过对多个分类器的分类结果进行某种组合来决定最终的分类,而这种方法在一定程度上可以提高模型的准确率和训练速度.从数据分析看,决策树模型的准确率和召回率要稍高于支持向量机模型,这是由于决策树具有在相对短的时间内能够对数据做出可行且效果良好的结果且对缺失值不敏感的特性.

为验证训练模型的准确率,我们从 2012—2015 年和 2017 年 1—7 月间中国地震台网记录到的 1058 次震动事件(含 4887 条台站纪录)中除去训练集部分,然后随机抽取天然地震事件 272 次(含 785 条台站波形),非天然震动事件 326 次(含 437 条台站波形),送入训练好的系统中进行自动分类,识别准确率为 86.62%. 具体事件识别结果如表 5 所示(训练集事件分布见图 1).

表 5 数据表明,本方法对内蒙、江苏、西藏、吉林、广东、重庆、贵州、黑龙江地区识别准确率小于 85%,效果还需加强,这可能是由于受到不同区域地质构造不同、地下速度结构不同、台站分布密度不同,随机抽取到的事件类型不同等因素影响而导致的.除此之外,其余地区的识别准确率均大于 85%,能够较为准确的对天然地震和非天然震动事件进行区分.

6 结论

本文使用 Bagging 采样策略,以决策树为基分类器,运用多数投票方法进行集成学习,在地震事件数据集上仿真实验的结果证明该方案是可行有效的.从实验和地震事件识别分析可知,该方法具有以下特点:一是通过 Bagging 采样策略,形成不同的训

表 5 震动事件预测结果

Table 5 Provincial and shock event predictions

所在省份	天然地震事件个数	非天然地震事件个数	判断对的天然地震个数	判断对的非天然地震个数	识别准确率 (%)
四川	103	14	96	13	93.16
云南	16	1	14	1	88.23
新疆	31	10	28	8	87.80
内蒙	11	77	9	59	77.27
江苏	5	6	2	6	72.72
西藏	4	0	3	0	75.00
辽宁	9	18	8	17	92.59
河北	7	9	7	9	100.00
甘肃	19	15	17	12	85.29
湖北	5	3	5	3	100.00
青海	7	2	6	2	88.88
吉林	8	13	6	9	71.42
陕西	3	31	2	29	91.17
福建	0	0	0	0	0.00
广东	10	2	7	2	75
山西	5	46	4	42	86.27
宁夏	4	5	4	4	88.88
重庆	4	1	3	1	80
贵州	3	0	2	0	66.67
天津	2	0	2	0	100.00
江西	3	0	3	0	100.00
山东	4	36	3	30	82.50
浙江	5	0	5	0	100.00
河南	2	14	2	10	75.00
台湾	0	0	0	0	0.00
广西	0	0	0	0	0.00
北京	0	14	0	14	100.00
黑龙	0	6	0	4	66.67
安徽	0	3	0	3	100.00
湖南	0	0	0	0	0
海南	0	1	0	1	100.00

练子集,可以在一定程度上缓解样本不足的问题.二是将决策树算法集成到一起,可以缓解单个分类器过拟合的问题.三是通过与非集成实验对比可知, Bagging 集成学习算法具有更快的训练速度和较好的稳定性.该方法可用于对地震事件类型进行初步的识别区分,形成实用的技术系统,实现对天然地震和非天然震动事件的准实时判定,从而实现震动事件类型的快速和准确判定;还可以在此基础上,进一

步将非天然震动事件进行分类,确认其具体事件性质(爆破、塌陷、滑坡等)。

未来可在此系统基础上,形成相对准确的中国地震台网统一地震目录和中国地震台网非天然震动事件目录,为后续的科研和紧急事件处理提供有效服务。

References

Allen R V. 1978. Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5): 1521-1532.

Allen R V. 1982. Automatic phase pickers: their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6B): S225-S242.

Baer M, Kradolfer U. 1987. An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, 77(4): 1437-1445.

Bai C, Huang L, Pan X, et al. 2018. Optimization of deep convolutional neural network for large scale image retrieval. *Neurocomputing*, 303: 60-67.

Bayar B, Stamm M C. 2018. Constrained convolutional neural networks: a new approach towards general purpose image manipulation detection. *IEEE Transactions on Information Forensics and Security*, 13(11): 2691-2706.

Bi M X, Huang H M, Bian Y J, et al. 2011. A study on seismic signal HHT features extraction and SVM recognition of earthquake and explosion. *Progress in Geophysics* (in Chinese), 26(4): 1157-1164, doi: 10.3969/j.issn.1004-2903.2011.04.004.

Bian Y J. 2002. Application of genetic BP network to discriminating earthquakes and explosions. *Acta Seismologica Sinica* (in Chinese), 24(5): 516-524.

Breiman L. 1996. Bagging predictors. *Machine Learning*, 24(2): 123-140.

Freund Y, Schapire R E. 1996. Experiments with a new boosting algorithm. // Proceedings of the 13th International Conference on Machine Learning. Bari, Italy: Morgan Kaufmann Publishers Inc., 148-156.

Giudicepietro F, Esposito A M, Ricciolino P. 2017. Fast discrimination of local earthquakes using a neural approach. *Seismological Research Letters*, 88(4): 1089-1096.

Han S Q, Song Z B, Wu H J. 2011. The application of improved algorithms of P-wave complexity in nuclear explosion monitoring. *Journal of Vibration and Shock* (in Chinese), 30(2): 205-209.

He X S, Li S Y, Shen P, et al. 2006. A wavelet packet approach to wave classification of earthquakes and mining shocks. *Earthquake Research in China* (in Chinese), 22(4): 425-434.

Huang H M, Bian Y J, Lu S J, et al. 2010. A wavelet feature

research on seismic waveforms of earthquakes and explosions. *Acta Seismologica Sinica* (in Chinese), 32(3): 270-276.

Kang Y, Chen L F, Lin W, et al. 2001. Identification of earthquake and blasting in Xinfengjiang. *South China Journal of Seismology* (in Chinese), 21(S1): 40-46.

Liu H, Zhang J Z. 2014. STA/LTA algorithm analysis and improvement of Microseismic signal automatic detection. *Progress in Geophysics* (in Chinese), 29(4): 1708-1714, doi: 10.6038/pg20140429.

Liu X Q, Shen P, Zhang L, et al. 2003. Using method of energy linearity in wavelet transform to distinguish explosion or collapse from nature earthquake. *Northwestern Seismological Journal* (in Chinese), 25(3): 204-209.

Long F, Ruan X. 2017. Challenges and research prospects of induced earthquakes. *Recent Developments in World Seismology* (in Chinese), (5): 11-15.

Maeda N. 1985. A method for reading and checking phase time in auto-processing system of seismic wave data. *Zisin*, 38(3): 365-379.

Melkl G, Cano A, Ventura S. 2018. MIRSVM: Multi-instance support vector machine with bag representatives. *Pattern Recognition*, 79: 228-241.

Perol T, Gharbl M, Denolle M. 2018. Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2): e1700578.

Pham B T, Bul D T, Dholakla M B, et al. 2016. A novel ensemble classifier of rotation forest and Naive Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. *Geomatics, Natural Hazards & Risk*, 8(2): 649-671.

Taylor S R. 2011. Statistical discriminants from two-dimensional grids of regional P/S spectral ratios. *Bulletin of the Seismological Society of America*, 101(4): 1584-1589.

Wang T T, Bian Y J. 2011. Criterion selection of earthquake and explosion recognition. *Seismological and Geomagnetic Observation and Research* (in Chinese), 32(6): 62-67.

Yang X H, Shen P, Liu X Q, et al. 2005. Application of method of spectral component ratio of wavelet-packets to discrimination between earthquakes and nuclear explosions. *Chinese Journal of Geophysics* (in Chinese), 48(1): 148-156.

Zeng R S, Ding Z F, Wu Q J, et al. 2000. Seismological evidences for the multiple incomplete crustal subductions in Himalaya and southern Tibet. *Chinese Journal of Geophysics* (in Chinese), 43(6): 780-797.

附中文参考文献

毕明霞, 黄汉明, 边银菊等. 2011. 天然地震与人工爆破波形信号 HHT 特征提取和 SVM 识别研究. *地球物理学进展*, 26(4): 1157-1164, doi: 10.3969/j.issn.1004-2903.2011.04.004.

边银菊. 2002. 遗传 BP 网络在地震和爆破识别中的应用. *地震学报*, 24(5): 516-524.

韩绍卿, 宋仔标, 伍海军. 2011. 改进的波形复杂度算法在核爆炸

监测中的应用. 振动与冲击, 30(2): 205-209.

和雪松, 李世愚, 沈萍等. 2006. 用小波包识别地震和矿震. 中国地震, 22(4): 425-434.

黄汉明, 边银菊, 卢世军等. 2010. 天然地震与人工爆破的波形小波特征研究. 地震学报, 32(3): 270-276.

康英, 陈丽芬, 林伟等. 2001. 新丰江地震与爆破的识别. 华南地震, 21(增刊): 40-46.

刘晗, 张建中. 2014. 微震信号自动检测的 STA/LTA 算法及其改进分析. 地球物理学进展, 29(4): 1708-1714, doi: 10.6038/pg20140429.

刘希强, 沈萍, 张玲等. 2003. 用小波变换能量线性度方法识别天然地震与爆破或塌方. 西北地震学报, 25(3): 204-209.

龙锋, 阮祥. 2017. 诱发地震带来的挑战及研究前景. 国际地震动态, (5): 11-15.

王婷婷, 边银菊. 2011. 识别天然地震和人工爆破的判据选择. 地震地磁观测, 32(6): 62-67.

杨选辉, 沈萍, 刘希强等. 2005. 地震与核爆识别的小波包分量比方法. 地球物理学报, 48(1): 148-156.

曾融生, 丁志峰, 吴庆举等. 2000. 喜马拉雅及南藏的地壳俯冲带——地震学证据. 地球物理学报, 43(6): 780-797.

(本文编辑 胡素芳)