

黄汉明, 边银菊, 卢世军, 蒋正锋, 李锐. 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.

Huang Hanming, Bian Yinju, Lu Shijun, Jiang Zhengfeng, Li Rui. 2010. A wavelet feature research on seismic waveforms of earthquakes and explosions. *Acta Seismologica Sinica*, 32(3): 270-276.

天然地震与人工爆破的波形小波特征研究^{*}

黄汉明^{1,†} 边银菊²⁾ 卢世军¹⁾ 蒋正锋¹⁾ 李锐¹⁾

1) 中国广西桂林 541004 广西师范大学计算机科学与信息工程学院

2) 中国北京 100081 中国地震局地球物理研究所

摘要 研究了如何从天然地震和人工爆破事件的波形记录中提取出有效、适用的波形特征, 以用于对爆破事件的识别. 首先对波形记录进行了4层小波包变换; 然后对变换得到的最后一层小波包系数提取3种波形特征: 能量比特征、香农熵特征及对数能量熵特征; 最后利用 ν -SVC支持向量分类机对这3种特征的分类能力进行了外推检验. 通过选用不同地区、不同台站、不同震级的天然地震与人工爆破的波形记录, 力求提取的特征量能尽可能地反映天然地震与人工爆破波形的本质区别, 尽量弱化震中距、震级等因素对识别效果的影响. 结果表明, 上述3种特征中以香农熵特征的识别效果最好, 能反映天然地震与人工爆破的本质区别, 可作为识别天然地震与人工爆破的一个有效判据.

关键词 爆破识别 小波包 能量比 香农熵 支持向量机

doi:10.3969/j.issn.0253-3782.2010.03.002 中图分类号: P315.63 文献标识码: A

A wavelet feature research on seismic waveforms of earthquakes and explosions

Huang Hanming^{1,†} Bian Yinju²⁾ Lu Shijun¹⁾ Jiang Zhengfeng¹⁾ Li Rui¹⁾

1) College of Computer Science and Information Engineering, Guangxi Normal University, Guangxi Guilin 541004, China

2) Institute of Geophysics, China Earthquake Administration, Beijing 100081, China

Abstract: Research on how to extract seismic wave features from earthquakes and explosions and how to discriminate explosions from earthquakes based on these features. Firstly, the transform of 4-layer wavelet packet is performed on the wave records. Secondly, the last layer coefficients of wavelet packet from the transform are employed to extract 3 types of wave features: energy ratio, Shannon entropy and logarithmic energy entropy. Thirdly, these features are supplied to a classifier of ν -SVC support vector machines for verifying the capabilities of these features. In order to weaken undesirable effect of event epicenter-distance and magnitude on the recognition, we tried to extract more essential features of the wave records gathered from different regions, different observa-

^{*} 基金项目 中国地震局地震行业科研专项基金(200808003)资助.

收稿日期 2009-06-05 收到初稿, 2009-11-04 决定采用修改稿.

[†] 通讯作者 e-mail: hhmmentor@gmail.com

tories and various events almost covering whole magnitude ranges. The results show that, among the above three features, the feature of Shannon entropy is the best candidate for discriminating explosions from earthquakes. This may be an effective criterion in explosion recognition.

Key words: explosion recognition; wavelet packet; energy ratio; Shannon entropy; v -SVC

引言

天然地震与人工爆破的震源性质是不同的. 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); 而人工爆破(包括化学爆破、地下核爆炸等)多位于地表附近, 其爆破源为对称膨胀源. 地震观测台网记录到的事件信号中包含了很多人工地震, 如矿山爆破等. 这些事件的记录, 如果不能及时剔除, 会混淆我们的地震目录, 影响地震学的研究工作.

在天然地震与人工爆破事件的分类识别中, 如何提取出有效的识别特征是识别的关键. 自上世纪 50 年代开始, 国内外在这两类事件的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、震源深度、体波震级 m_b 与面波震级 M_s 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、瞬态谱等(刘希强等, 2003; 边银菊, 2002; 杨选辉等, 2005; 和雪松等, 2006; Tjøstheim, 1978). 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, 某些判据仅仅适用于一些特定的地区和台站记录. 本文试图从波形记录中提取出有效、适用的波形特征, 以用于对上述两类事件的识别.

1 小波包分析方法基本原理

1.1 小波包分析

小波包分析是小波变换中的一种. 小波包对信号的低频近似部分及高频细节部分进行逐层分解, 可以把信号按频带分得更细, 这样就可以了解到信号中包含的更多细节信息. 小波包由一簇正交紧支小波构成, 可以用如下函数簇描述(高志, 余啸海, 2007):

定义小波包(wavelet packet, WP): 若子空间 U_j^n 是函数 $u_n(t)$ 的闭包空间, U_j^{2n} 是函数 $u_{2n}(t)$ 的闭包空间, 并令 $u_n(t)$ 满足下面的双尺度方程:

$$\begin{cases} u_{2n}(t) = \sqrt{2} \sum_{k \in Z} h(k) u_n(2t - k) \\ u_{2n+1}(t) = \sqrt{2} \sum_{k \in Z} g(k) u_n(2t - k) \end{cases} \quad (1)$$

式中, $h(k)$, $g(k)$ 为正交镜像滤波器组, 满足 $g(k) = (-1)^k h(1-k)$. 当 $n=0$ 时, 式(1)直接给出

$$\begin{cases} u_0(t) = \sum_{k \in Z} h(k) u_0(2t - k) = \phi(t) \\ u_1(t) = \sum_{k \in Z} g(k) u_0(2t - k) = \varphi(t) \end{cases} \quad (2)$$

式中, $\phi(t)$ 和 $\varphi(t)$ 分别为多分辨分析中的尺度函数与小波基函数. 我们称由式(2)构造的序

列 $\{u_n(t)\}$, $n \in Z_+$ 为由 $u_0(t) = \phi(t)$ 确定的正交小波包.

对实际地震信号(2001-10-12T11:28:49.3, 38.46°N、113.06°E, $M_2.3$, $h=33.0$ km)进行小波包4层分解后得到16个子频带信号的波形如图1所示. 这16个子频带信号的频带分布是从第1个WPT系数到第16个WPT系数顺序增加的. 对地震和爆破信号, 只选择各台站BHE分量, 采样率为50 Hz. 样本波形分解为16个子频段后, 很多样本的前7个子频段对应的特征向量有值, 后面的几乎全为零. 从地震、爆破两类样品的各子频段的频谱图上获得的平均能量和卓越频率分别做柱状分布图, 地震和爆破有些差异, 但难以根据任何一个子频段的特性来较可信地区分出地震与爆破. 故需要根据一定的算法综合利用这些单个来看并不明显的子频段特征. 为此本文研究了香农熵等3个综合性特征.

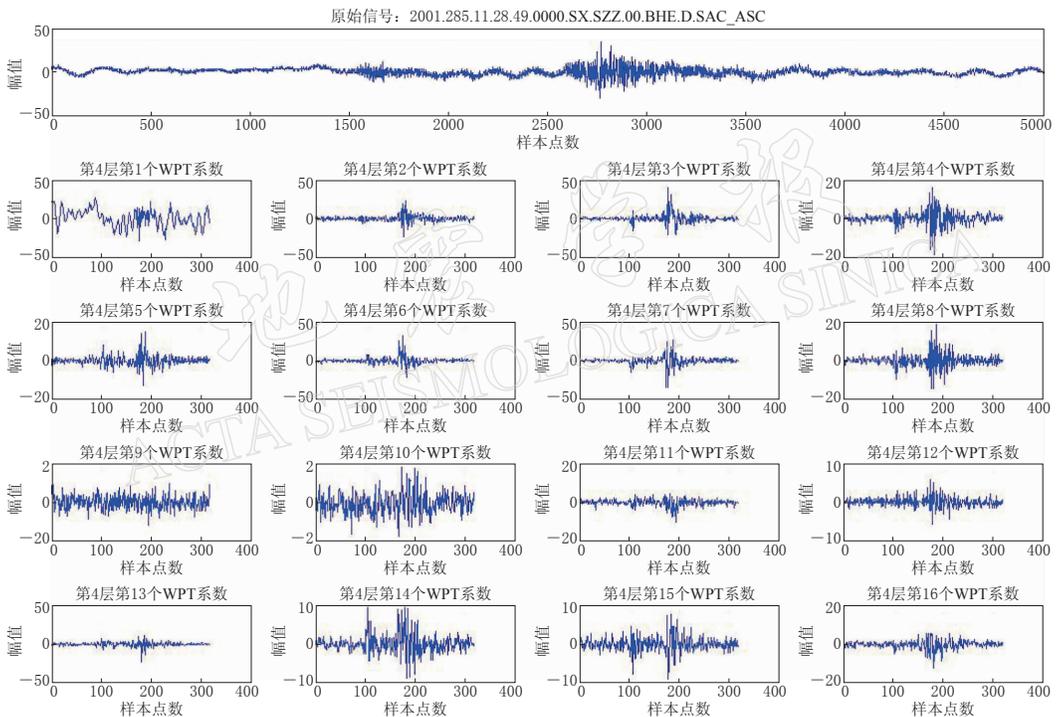


图1 天然地震信号的4层小波包分解

Fig. 1 4-layer wavelet packets decomposition of a seismic signal

1.2 本文所选取的波形特征

本文对各天然地震与人工爆破记录进行4层小波包分解后, 对第4层的小波包系数分别提取出3种波形特征(均为16维向量): 小波系数的能量比(E_{wt})、小波系数的香农熵($E_{shannon}$)和小波系数的对数能量熵(E_{lg}), 研究了各特征的分类能力.

若 S 为原始信号, 对它进行 n 层小波包分解后, 得到第 n 层的小波包系数总共为 N 个. S_i 为信号 S 分解后的第 i 个小波的系数, 其长度为 J , 小波系数的结点序号为 j , 则从第 i 个小波的系数中提取出的能量比(E_{wt})、香农熵($E_{shannon}$)和对数能量熵(E_{lg})特征分别按式(3)、(4)和(5)定义.

$$E_{\text{wt}}(i) = 100 \times \frac{\sum_{j=1}^J S(i, j)^2}{\sum_{n=1}^N \sum_{m=1}^J S(n, m)^2} \quad (3)$$

$$E_{\text{shannon}}(i) = - \sum_{j=1}^J S(i, j)^2 \lg[S(i, j)^2] \quad (4)$$

$$E_{\lg}(i) = \sum_{j=1}^J \lg[S(i, j)^2] \quad (5)$$

2 数据集的选取和特征提取

2.1 数据集的选取

为了最大限度地提取适用于不同地区、不同台站记录的波形特征,用于天然地震与人工爆破事件的识别,同时力求特征尽可能地反映两类事件的本质区别,尽量弱化震中距、震级等因素对识别效果的影响.本文在数据集的选取上有意识地选择了不同地区、不同台站、不同震级的天然地震与人工爆破事件的波形记录(对不同台站的同一次地震或人工爆破事件的记录当作不同的样本对待).另外只选择了各台站的 BHE 分量且有较长时间有效记录(振幅明显且记录中未显示出仪器有零点漂移)的波形数据来进行实验.

实验中所用的天然地震记录样本来自于从中国地震台网中心地震数据管理与服务系统(<http://www.csdmc.ac.cn/newweb/>)与国家地震数据共享中心(<http://data.earthquake.cn/data>)中下载的 SEED 文件中提取出来的 1336 条地震波形记录文件集合,其中包括:① 2001—2007 年以(39.44°N, 112.42°E)、(26.07°N, 101.81°E)、(25.16°N, 113.68°E)、(40.80°N, 116.96°E)和(31.30°N, 101.81°E)为中心,±1°区域范围内震级在 1.3—4.0 之间的地震波形记录 330 条;② 2001—2003 年发生于我国及周边地区震级在 5.3—7.9 之间的 303 条地震波形记录;③ 2008 年 5 月四川汶川地震的余震记录 703 条.

人工爆破记录样本集中总共有 236 条波形记录文件,其中包括:① 1987—1988 年记录到我国新疆及周边地区的人工爆破波形记录 31 条;② 2007 年宁夏爆破和延庆爆破的波形记录 205 条.

实验中从上面所述的天然地震和人工爆破数据集中,各随机选择 100 条记录总计 200 条记录用于实验(其中测试集和训练集均各有 100 条记录:其中的 50 条为天然地震记录,另 50 条为人工爆破记录).按如上方式选择中国及周边区域中几个典型区域的地震和爆破数字波形记录,目的是想突出震源特性在波形特征中的表现,试图为对中国及周边区域中的地震与爆破识别提供较好的基于震源特性的识别特征.

2.2 特征提取

不同事件或不同台站记录的整个地震波形的记录长度会有差异,有时有效记录长度差别很大.而特征提取时应针对同样长度的波形记录,本文采用某一长度的矩形窗对长度不一的原波形记录进行截取.采用不同的窗长度进行截取时,对所提取出来的波形特征的识别效果是有影响的.对于本研究所选用的波形记录,尤其是爆破记录,由于对应的震级很小,没有考虑从 P 波、S 波的角度截取波段.为了简化波段截取且试图可应用于任何震源、震级的波形,本文选择了以波形幅值绝对值最大的采样点为中心、窗长度为 500—3000,

按 500 为增量, 共 6 种长度的窗长度, 以及整个波形记录信号进行实验比较. 实验结果显示, 这 7 种长度的波形记录中, 窗长度为 2000 点时, 识别效果是可以接受的, 且是最好的. 所以, 本文的特征提取是对以波形最大绝对幅值采样点为中心的窗长度为 2000 的波段进行的.

本文中利用下述方法来提取各样本的波形特征:

1) 将各训练集(或测试集)样本用 db7 小波基函数进行 4 层小波包分解.

2) 对第 4 层的 16 个小波包系数分别利用公式(3)、公式(4) 和公式(5)提取各自的相应波形特征, 并将它们分别组合成一个 16 维的特征向量.

3) 利用标准差标准化方法对特征向量集进行标准化, 得到训练集(或测试集)的特征向量矩阵, 以此作为支持向量机分类器的输入参数, 利用下面(2.3 节)所述的参数所构造的分类器来检验各波形特征的分类能力.

2.3 分类器参数选择

为了研究经过小波变换后提取出来的波形特征在识别两类事件中的分类能力, 本文利用 Matlab 对此进行了实验验证, 分类器选择了 ν -支持向量分类机(ν -support vector classifier, ν -SVC) (郭丽娟等, 2008; 王睿, 2007). 具体参数设置如下:

1) ν -SVC 中的 $\nu=0.5$.

2) 核函数采用 Sigmoid 核: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh[\gamma(\mathbf{x}_i \cdot \mathbf{x}_j) + c]$, 其中, $\mathbf{x}_i, \mathbf{x}_j$ 分别为第 i 和第 j 个样品向量. 所谓样品向量, 在此即为前一小节(2.2 节)中所述的波形特征向量.

3) Sigmoid 核中的 $\gamma=1/100, c=0$.

2.4 支持向量机分类器的训练

将标准化后的训练集的特征向量利用文献(郭丽娟等, 2008; 王睿, 2007)中的支持向量机算法进行训练后, 得到分类器. 分类器模式的相关参数包括: 两类训练集中支持向量的数目和具体记录个数、各支持向量的对应权值、支持向量机算法中的阈值等. 表 1 为用训练得到的分类器对训练集本身进行识别的结果.

表 1 ν -SVC 对训练集本身的分类效果

Table 1 Classification result of training sets by ν -SVC

波形特征	分类能力检验结果				分类准确率
	天然地震		人工爆破		
	训练样本总数	误识数	训练样本总数	误识数	
能量比特特征	50	4	50	10	86%
香农熵特征	50	0	50	3	97%
对数能量熵特征	50	19	50	23	58%

3 分类能力检验结果

将各测试集样本用 2.2 节同样的方法进行特征提取, 用训练得到的支持向量分类机进行地震事件的分类, 以检验各波形特征的分类能力. 实验结果如表 2 及图 2 所示.

图 2 所对应的 ν -SVC 的决策函数(分类器)为

$$f(\mathbf{x}) = \operatorname{sgn}\left\{\sum_{i=1}^{51} y_i \times a_i^* \times \tanh[(\mathbf{x}_i \cdot \mathbf{x})/100] - 0.3558\right\} \quad (6)$$

式中, x_i 为训练集经支持向量机训练后确定的支持向量($i=1, \dots, 51$), y_i 等于 1(当第 i 个支持向量是天然地震记录时)或 -1(当第 i 个支持向量是人工爆破记录时), a_i^* 为第 i 个支持向量对应的权值, x 为待识别的测试集记录, -0.3558 为支持向量机训练后确定的阈值.

表 2 ν -SVC 对测试集的分类效果Table 2 Classification results of testing sets by ν -SVC

波形特征	分类能力检验结果				分类准确率
	天然地震		人工爆破		
	训练样本总数	误识数	训练样本总数	误识数	
能量比特征	50	9	50	11	80%
香农熵特征	50	1	50	3	96%
对数能量熵特征	50	31	50	23	46%

4 讨论与结论

1) 国内外利用其它判据进行地震与爆破的识别研究中, 准确率较高的可达到 90% 以上. 在本研究的 3 种波形特征中, 香农熵特征的分类能力达到了 97%, 用随机选择的测试集进行检验的识别效能也达到了 96% 的准确率, 识别效果较好. 这表明了香农熵特征能反映天然地震与人工爆破的本质区别, 可以作为识别天然地震与人工爆破的一个有效判据; 而能量比特征差一些, 还需进一步研究; 对数能量熵特征则很差, 可淘汰.

2) 相关文献中一般是对某一地区、在一定震级范围内的地震与爆破进行识别, 本文在较大的地区、较大的台站距离范围及较大的震级范围内研究了地震与爆破的识别, 这是我们为得到一个较普遍适用的识别判据进行的初步尝试, 实验结果显示应该是可行的. 此外, 为得到一个较普适的判据, 由于事件的震级范围跨度较大, 为避免震级相差较大带来的影响, 在提取特征前对波形数据还做了幅值归一化处理.

当然, 本文得到的香农熵特征的识别效能及普遍适用性还须在实践中经大量事件的验证.

3) 天然地震与人工爆破本质的差异是震源性质, 其次是传播路径的差异. 本文选择中国及周边区域中的几个典型区域、较大的台站距离范围及较大的震级范围研究地震与爆破的识别, 目的是想突出震源性质在波形特征中的表现, 对中国及周边区域中的地震与爆破识别提供较好的识别特征. 本文得到的香农熵特征的识别率为 96%, 表明结果较好. 但并未达到 100% 的识别, 今后应进一步考虑地震波传播路径差异的影响和波形记录中出现的多种震相种类及其先后次序的影响, 那么识别率应该更高, 结论会更可靠.

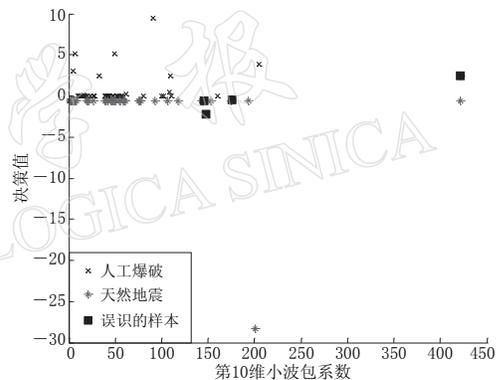


图 2 香农熵特征识别效果

Fig. 2 Recognition result based on Shannon entropy features

参 考 文 献

- 边银菊. 2002. 遗传 BP 网络在地震和爆破识别中的应用[J]. 地震学报, **24**(5): 516-524.
- 高志, 余啸海(编著). 2007. Matlab 小波分析与应用[M]. 北京: 国防工业出版社: 1-150.
- 郭丽娟, 孙世宇, 段修生. 2008. 支持向量机及核函数研究[J]. 科学技术与工程, **8**(2): 487-490.
- 和雪松, 李世愚, 沈萍, 冯全雄. 2006. 用小波包识别地震和矿震[J]. 中国地震, **22**(4): 425-434.
- 刘希强, 沈萍, 张玲, 李永红. 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[J]. 西北地震学报, **25**(3): 204-209.
- 王睿. 2007. 关于支持向量机参数选择方法分析[J]. 重庆师范大学学报(自然科学版), **24**(2): 36-42.
- 杨选辉, 沈萍, 刘希强, 郑治真. 2005. 地震与核爆识别的小波包分量比方法 [J]. 地球物理学报, **48**(1): 148-156.
- 曾融生, 陈运泰. 2000. 探测地球内部的“雷达”: 地震波(续)[J]. 城市防震减灾, (6): 12-14.
- Tjøstheim D. 1978. Improved seismic discrimination using pattern recognition [J]. *Phys Earth Planet Inter*, **16**(2): 85-108.

地 震 学 报
ACTA SEISMOLOGICA SINICA