王闯,刘伟,靳源,等.基于Bitmap的地震波形数据连续性评价方法及应用[J].地震工程学报,2026,48(5):1189-1196.DOI:10.20000/j.1000-0844.20240410001

WANG Chuang, LIU Wei, JIN Yuan, et al. Bitmap-based method for assessing seismic waveform data continuity and its application[J]. China Earthquake Engineering Journal, 2026, 48(5): 1189-1196. DOI: 10.20000/j.1000-0844.20240410001

基于Bitmap的地震波形数据连续性评价方法及应用

王 闯,刘 伟,靳 源,周 聪,柴旭超,王文青

(中国地震局第二监测中心,陕西 西安 710054)

摘要:数据连续性是衡量地震波形数据质量的重要评价指标。针对现有连续性评价方法在应用于分段数过多或分段间异常排列等实际数据场景时存在的耗时长、算法失效等问题,提出一种基于Bitmap的波形数据连续性评价方法。通过对原始波形构建其Bitmap位向量,将波形数据采样点映射为表示记录状态和重叠状态的二进制位,直观反映评价时间窗口内数据记录的连续性,从数据结构上消除分段乱序、重叠所带来的影响,从而快速得到可靠的指标结果。在真实应用场景下的实验结果表明,该方法具有明显的性能优势和更强的鲁棒性,对于开展中国大陆测震台网数据连续性评价具有现实意义。

关键词:地震波形数据;数据连续性评价;数据质量;Bitmap

中图分类号:P315 **文献标志码**:A **文章编号**:1000-0844(2026)05-1189-08

DOI:10.20000/j.1000-0844.20240410001

Bitmap-based method for assessing seismic waveform data continuity and its application

WANG Chuang, LIU Wei, JIN Yuan, ZHOU Cong, CHAI Xuchao, WANG Wenqing

(*The Second Monitoring and Application Center, CEA, Xi'an 710054, Shaanxi, China*)

Abstract: Data continuity is a crucial metric for evaluating the quality of seismic waveform data. Current methods for assessing seismic waveform data continuity often face challenges such as excessive file segments or irregular arrangements among data segments, leading to significant time consumption or algorithm failure. To address these issues, this study introduces a bitmap-based method for evaluating waveform data continuity. This method constructs bitmap vectors to represent the recording and overlapping statuses by mapping data sampling points into bits. The constructed bitmap vectors effectively capture the continuity of data recording within the assessment time window and eliminate the impact of disordered or overlapped data segments from a data structure perspective. This approach

收稿日期:2024-04-10

基金项目:地震科技星火计划(XH23060YA);陕西省自然科学基础研究计划(2024JC-YBQN-0332)

第一作者简介:王 闯(1989-),男,硕士,高级工程师,主要从事地震监测数据质量评价和地震信息化工作。E-mail:wch_xa@163.com。

enables rapid and accurate calculation of continuity metrics. Experimental results in real-world scenarios demonstrate the method's superior performance, robustness, and practical value for conducting reliable continuity assessments on waveform data from the China seismic network.

Keywords: seismic waveform data; data continuity assessment; data quality; Bitmap

0 引 言

地震波形数据质量直接关系到地震学研究结果的可信度,对其质量进行及时、有效的评价具有重要意义[1-2]。数据连续性是衡量波形数据质量最为基础的一项指标[3-4],其反映了连续观测数据在整个时间序列上的完整性[5-6],是国内外数据中心最早纳入数据质量监控的指标之一[1]。连续性良好的地震波形数据是地震学研究的可靠基础,也是保障监测预警业务正常运行的前提[7]。导致数据连续性不佳的原因主要是仪器故障、供电、通信等问题[8],这些问题会造成数据的大片缺失或频繁断记,严重影响数据的可用性。

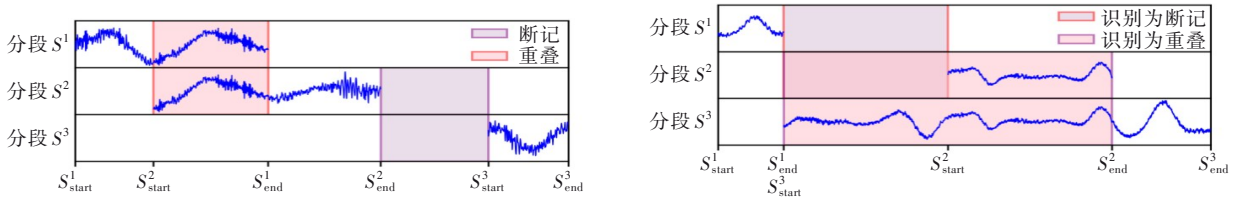
地震波形数据连续性的评价方法主要有两类:一种是抽样调查法,即发起大量随机数据请求,每个请求仅获取数秒或数分钟的数据,然后通过对数据获取的成功率进行统计分析来间接评价连续性。这种方法无需获取全部波形数据,效率较高,但结果的精确性有限。另一种方法是获取评价时间段内的全量数据,基于此构建断记列表来完成连续性评价[9]。目前主流的地震波形数据质量评价软件,如美国地震学联合会数据中心的MUSTANG[10]、美国地质调查局阿尔伯克基地震实验室的DQA[11]、美国桑迪亚国家实验室的Pycheron[12]等,均采用第二种方法。这些软件通过计算完整率、断记次数等指标定量评价数据的连续性,但在处理实际数据中常见的分段过多或乱序等异常情况时,往往存在耗时过长甚至算法失效的问题,影响了评价结果的时效性和可靠性。

为解决现有评价方法在性能和鲁棒性上的不足,本文提出一种基于Bitmap的地震波形数据连续性评价方法。该方法通过构造数据采样点的二进制位向量来映射记录状态,从数据结构上消除分段异常带来的负面影响,从而能够快速、可靠地计算出评价指标。本文通过在仿真数据和真实数据上与现有方法进行对比实验,验证了所提方法的有效性和实用性。目前,该方法已应用于中国大陆测震台网的日常数据质量评价工作。

1 现有评价方法

现有主流波形数据质量评价软件所采用的连续性指标算法均基于波形数据的存储特点进行设计,评价对象通常是按日归档存储的波形数据文件。目前应用最为广泛的波形数据质量评价软件是ISPAQ[13-16],其对数据连续性的评价方法在同类软件中具有代表性:首先,获取被评价数据中的数据分段列表;然后,记录相邻分段间的间隔,间隔长度为每一分段起始时间与上一分段结束时间之差,将正间隔记为断记并构建断记列表,将负间隔记为重叠并构建重叠列表;最后,根据断记和重叠详情计算完整率、断记次数、重叠次数等评价指标。该方法实现了对波形数据连续性的定量评价,但对数据规范性有一定要求,数据分段数量和分段间的排列顺序均可能对算法性能产生影响。

图1(a)给出了一个同时包含断记和重叠的典型场景,3个分段S^1、S^2、S^3按顺序排列,其中S^i_{start}和S^i_{end}分别表示分段S^i的起始时间和结束时间。显然,S^1、S^2间的间隔长度$S^2_{\text{start}}-S^1_{\text{end}}$为负,$S^2$、$S^3$间的间隔长度$S^3_{\text{start}}-S^2_{\text{end}}$为正,现有方法在该场景下能够有效识别重叠和断记现象。图1(b)则给出了一例因分段乱序引起的算法失效场景:S^1、S^2间的间隔长度$S^2_{\text{start}}-S^1_{\text{end}}$为正,被记为断记;$S^2$、$S^3$间的间隔长度$S^3_{\text{start}}-S^2_{\text{end}}$为负,被记为重叠。但该段数据事实上并未发生断记,且重叠也仅发生在S^2的时间范围内。现有方法在这种情况下出现了算法失效,且由于分段间存在相互包含关系,即使对数据分段重新进行排序仍不能改善算法效果,无法得出正确的断记和重叠列表。在实际的波形数据汇集中,数据分段较多、分段异常排列等情况时有发生[17]。这些非常规情况会为鲁棒性较差的评价算法或设计不完备的评价软件带来诸如计算耗时增加、算法失效等负面影响,进而影响评价结果获取的及时性和可靠性,而结果可靠性的损失是数据质量评价工作所不能接受的。



(a) 含有断记和重叠现象的典型场景　　(b) 存在分段乱序的场景

图 1　现有连续性评价方法在不同场景下的表现

Fig. 1　Performance of current methods for assessing seismic data continuity in different scenarios

2　基于 Bitmap 的评价方法

Bitmap 是一种适合大规模数据统计的数据结构[18]。其采用二进制位对数据元素的状态进行标记，将数据对象的索引以二进制位向量的形式存储，具有空间开销较小、查询响应速度快的优势[19-20]。Bitmap 尤其适用于分布稠密的数据，而地震波形数据作为时间序列数据，具有显著的稠密特征。

为消除断记、乱序等情况带来的负面影响，提升波形数据连续性评价的性能，本文提出一种基于 Bitmap 的波形数据连续性评价方法。该方法包括 2 个阶段：第 1 阶段，根据待评价波形数据构建 Bitmap 位向量；第 2 阶段，基于该位向量完成各项指标的计算。

2.1　Bitmap 位向量构建

根据待评价波形构建 Bitmap 位向量是本方法的关键步骤。每条待评价波形数据需构建 2 个 Bitmap 位向量：记录位向量 $\boldsymbol{V}$ 和重叠位向量 $\boldsymbol{U}$，分别用于表示评价窗口内各采样时刻的记录状态和重叠状态。位向量中的每个二进制位与评价窗口中期望记录到的每个采样点相对应，位向量的长度与评价窗口内期望的采样点数量一致。要构建的位向量与待评价数据的映射关系为：

$$V^i=\begin{cases}1, & \text{if } Sample[i] \text{被记录}\\ 0, & \text{else}\end{cases}$$
$$U^i=\begin{cases}1, & \text{if } Sample[i] \text{出现重叠}\\ 0, & \text{else}\end{cases} \tag{1}$$

式中：$Sample$ 表示波形数据采样点；i 表示采样点索引；V^i 表示位向量 $\boldsymbol{V}$ 中的第 i 位；U^i 表示位向量 $\boldsymbol{U}$ 中的第 i 位。

构建位向量的过程如下：新建并初始化记录位向量 $\boldsymbol{V}$ 和重叠位向量 $\boldsymbol{U}$，向量长度与评价窗口内的期望采样点数相同。然后遍历待评价波形记录中的各数据分段 $\boldsymbol{S}$，根据数据分段的起始时间和结束时间确定当前分段在位向量中的起始和结束位置索引。此时，若 $\boldsymbol{V}$ 中对应位置的值已为 1，则表明之前记录的分段与当前分段存在重叠，需将 $\boldsymbol{V}$ 中对应位置的值复制到重叠位向量 $\boldsymbol{U}$ 中的相同位置，以指示数据重叠；然后，将记录位向量 $\boldsymbol{V}$ 中对应位置的二进制位设置为 1，以指示该时间段内有数据存在。遍历结束后，位向量构建完毕。具体算法如表 1 所列。

表 1　Bitmap 位向量构建算法

Table 1　Bitmap bit-vector construction algorithm

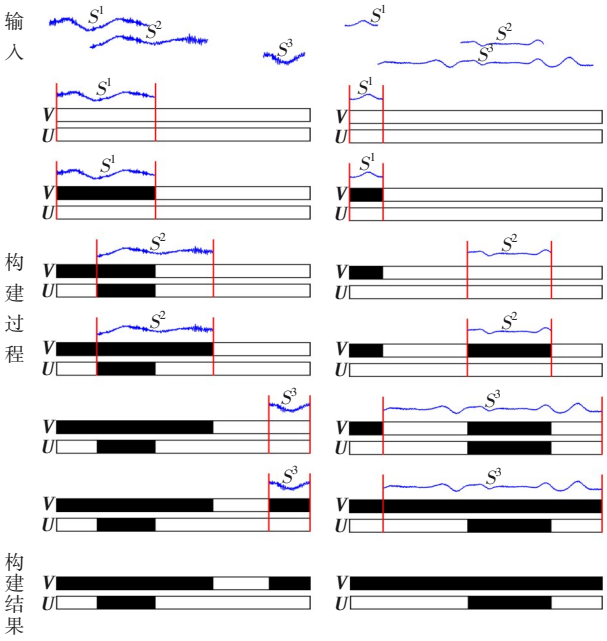
算法：Bitmap 位向量构建算法	
	输入：数据分段列表 $\boldsymbol{S}=\{S^1,\cdots,S^N\}$；采样率 F_s；评价窗口 (t_{start},t_{end})
	输出：记录位向量 $\boldsymbol{V}$；重叠位向量 $\boldsymbol{U}$
1	$L \leftarrow F_s\times(t_{end}-t_{start})$ // 要构建的位向量长度
2	$\boldsymbol{V}\leftarrow \text{zeros}(L)$, $\boldsymbol{U}\leftarrow \text{zeros}(L)$
3	for $i\leftarrow 1,\cdots,N$ do // 遍历数据分段
4	$j_s\leftarrow F_s\times\left[\max\left(S^i_{start},t_{start}\right)-t_{start}\right]$
5	$j_e\leftarrow F_s\times\left[\min\left(S^i_{end},t_{end}\right)-t_{start}\right]$
6	for $j\leftarrow j_s,\cdots,j_e$ do
7	$U^j\leftarrow V^j$
8	$V^j\leftarrow 1$
9	end for
10	end for
11	return $\boldsymbol{V}$, $\boldsymbol{U}$

注：N 为数据分段列表长度，即列表中的元素个数；L 为评价窗口内期望记录到的采样点数；zeros 函数表示创建指定长度的零向量；F_s 通常为 100 Hz；S^i_{start} 和 S^i_{end} 分别为分段 S^i 的起、止时间；t_{start} 和 t_{end} 分别是评价窗口起、止时间，单位为 s，窗口时长通常为一个自然日

图 2 绘出了构建算法的具体示例，包括输入波形、Bitmap 位向量构建的分步结果和最终映射结果，重点描述了对位向量的赋值过程。图 2(a) 对应图 1 中的典型数据场景，构建过程中按顺序遍历 3 个分段：首先读入 S^1，将 $\boldsymbol{V}$ 在 S^1 时间范围内的值复制到 $\boldsymbol{U}$ 中对应位置，由于此前 $\boldsymbol{V}$ 和 $\boldsymbol{U}$ 刚完成初始

化，U中的值未发生变化，然后将V在S^1时间范围内的值赋1以记录S^1；接着读入S^2，将V在S^2时间范围内的值复制到U中对应位置，由于此前该范围内V中已有部分位被赋1，U中对应位置也将置1，从而记录发生的重叠，然后将V在S^2时间范围内的值赋1以记录S^2；最后读入S^3，将V在S^3时间范围内的值复制到U中对应位置，由于此前该范围内V中值全为0，U中的值未发生变化。最终的映射结果中，可看到V和U清晰指示了数据的记录和重叠情况。图2(b)对应图1中的异常数据场景，构建过程与图2(a)一致。值得注意的是，在对S^3进行处理时，算法未受到分段间相互关系的影响，有效杜绝了分段乱序造成的问题。

构建的Bitmap位向量通过对采样点记录和重叠状态的逐点映射，从数据结构层面消除了分段数量过多、乱序等异常情况对计算速度与结果准确性的负面影响，从算法层面保证了计算结果的可靠性。位向量提供了一种简洁而有效的数据连续性表达形式，能够直观反映评价时间窗口内数据记录的连续状况。由于位向量精确对应到了每一个采样点，其对数据断记和重叠现象的分辨颗粒度可达采样间隔级别。



(a) 含有断记和重叠现象的典型场景　(b) 存在分段乱序的场景

图2 Bitmap位向量的构建

Fig. 2 Construction of Bitmap bit-vector

2.2 指标计算

基于构建的Bitmap位向量完成评价指标计算，常用的地震波形数据连续性评价指标及含义列于表2。

表2 波形数据连续性评价指标

Table 2 Metrics of waveform data continuity assessment

指标名称	指标含义
数据完整率(percent_availability)	表示24 h内实际记录到的数据量占期望记录数据量的比例，100%意味着当天数据没有缺失
数据分段数(num_segs)	表示24 h内的数据分段数量，其中数据分段指一段无中断的连续序列
断记次数(num_gaps)	表示24 h内数据断记的发生次数，在数据头部或尾部的数据缺失同样记为断记
最大断记时长(max_gap)	表示24 h内遇到的最大断记的时间长度，单位为s
重叠次数(num_overlap)	表示24 h内数据重叠现象的发生次数
最大重叠时长(max_overlap)	表示24 h内遇到的最大重叠的时间长度，单位为s

数据完整率：通过对记录状态位向量V进行遍历求和计算得出，其计算公式为：

$$percent_availability=\frac{\sum_i V^i}{L}\times 100\% \tag{2}$$

数据分段数：即从位向量V中提取出的数据分段列表S所包含的元素个数，计算公式为：

$$num_segs=\mathrm{count}(S)=N \tag{3}$$

式中：count函数用于计算列表中的元素数量。

断记次数和最大断记时长：首先，遍历搜索记录状态位向量V，将其中连续为0的二进制子串识别为一次断记，并记录其起始位置索引和持续长度，从而得到断记列表G。该过程可表示为：

$$G=\mathrm{find_consecutive}(V,0)=\{G^1,\cdots,G^K\} \tag{4}$$

式中：K为列表G中元素个数；find_consecutive函数实现了在位向量中搜索连续为0或1的二进制子串的功能，并返回所有子串的信息。每个断记子串记作G^i，其包含了该子串在位向量中的起始位置索引G^i_{start}和其持续长度G^i_{length}。据此，断记次数即断记列表G中的元素个数：

$$num_gaps=\mathrm{count}(G)=K \tag{5}$$

最大断记时长则为该列表中所有断记时长的

最大值：

$$max_gap = \max\left(G^{i}_{\text{length}}\right) \tag{6}$$

重叠次数和最大重叠时长：首先，遍历搜索重叠状态位向量$\boldsymbol{U}$，将其中所有连续为1的二进制子串识别为一次数据重叠，记录其起止位置索引和持续长度，得到重叠列表$\boldsymbol{O}$：

$$\boldsymbol{O} = \text{find_consecutive}(\boldsymbol{U}, 1) = \left\{O^{1}, \cdots, O^{M}\right\} \tag{7}$$

式中：M为列表$\boldsymbol{O}$中的元素个数；O^{i}表示一个搜索到的重叠子串，同样包含其位置索引O^{i}_{start}和持续长度O^{i}_{length}。重叠次数即为重叠列表$\boldsymbol{O}$中的元素个数：

$$num_overlaps = \text{count}(\boldsymbol{O}) = M \tag{8}$$

最大重叠时长则为该列表中所有重叠时长的最大值：

$$max_overlap = \max\left(O^{i}_{\text{length}}\right) \tag{9}$$

上述计算流程所得的各项指标，能够全面覆盖当前主流数据质量评价软件所支持的核心连续性评价指标。此外，本方法还能直接提供结构化的断记列表和重叠列表。利用这些列表中每个断记或重叠事件的起止时刻信息，可以反向重建位向量，从而能够更细致、更全面地刻画数据的连续性特征。

2.3　与现有评价方法的计算结果对比

为验证所提方法的有效性，分别使用本文提出的Bitmap方法与现有工具ISPAQ对图1中“典型数据场景”与“异常数据场景”下的仿真数据进行连续性指标计算并对比。仿真设置均以0时刻作为记录开始时间，采样率为100 Hz。其中，典型数据场景包含3个数据分段（S^{1}、S^{2}、S^{3}），其起止时间分别为0～7.83 s、3.17～12.38 s、16.75～20 s；异常数据场景同样包含S^{1}、S^{2}、S^{3}，其起止时间分别为0～3.99 s、14～23.99 s、4～30 s。计算结果显示，在典型数据场景下，两种方法均准确计算出数据完整率为78.2%，结果一致。然而，在异常数据场景下，ISPAQ计算出的数据完整率为66.67%，出现了误判；而Bitmap方法的计算结果为100%，有效避免了误判发生。这一结果与本文第1节中的分析一致，证明了Bitmap方法在处理分段乱序、重叠等异常情况时具有更高的准确性。

3　实验与讨论

3.1　实验环境与数据

实验硬件环境为单台服务器，配置为英特尔至强Gold 5220 2.20 GHz处理器，16 GB内存，操作系统为Ubuntu 18.04 LTS。

实验数据分为2组：数据集Ⅰ为仿真数据集，模拟了一组数据分段依次递增的波形记录。分段对应的断记次数从0到5 000次，以200次为步长递增，共生成26条记录，总数据量为547 MB。每条记录中的断记均匀分布，每次断记持续1 s（即100个采样点）。数据集Ⅱ为国家地震数据灾备中心（简称灾备中心）[21]在2023年实时备份的中国大陆测震台网真实数据，包含参与全国测震数据质量评价的925个宽频带台站记录，共计974 857个文件，总数据量为10.68 TB。所有文件的波形记录时长均为24 h，采样率为100 Hz。

3.2　仿真数据耗时测试

在数据集Ⅰ上，分别使用本文提出的Bitmap方法、ISPAQ及Pycheron计算连续性指标，并对比其计算耗时。从图3可以看出，Pycheron的计算耗时随数据分段数量的增加而显著上升，这是由于其底层依赖obspy工具包[22]中的get_gaps函数，该函数为处理分段间的包含关系而采用了计算复杂度较高的算法。ISPAQ在计算前需将数据转换为R语言对象，产生了较大的预处理开销。相比之下，Bitmap方法的计算耗时对数据分段数量的增长不敏感，显示出显著的性能优势。

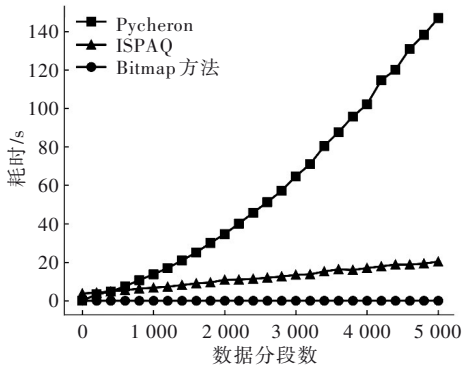


图3　不同算法计算连续性指标的耗时对比

Fig. 3　Comparison of computational time of different methods for assessing data continuity

3.3　在中国大陆测震台网真实数据上的应用

在数据集Ⅱ上，分别应用Bitmap方法、ISPAQ、Pycheron串行计算各台站波形数据的连续性指标，对比完成全年数据计算的总耗时。为确保对比的客观性，测试中排除了磁盘I/O时间的影响。图4展示了不同方法的累计计算耗时随日期（2023年）

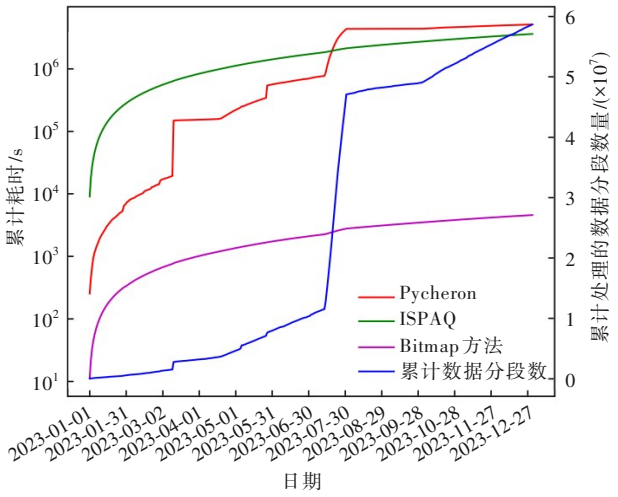


图4 数据集Ⅱ上不同算法计算连续性指标的累计耗时

Fig. 4 Comparison of cumulative computational time of different methods for assessing data continuity on dataset Ⅱ

的变化情况。从图中可以看出，Pycheron和ISPAQ的计算耗时明显受到数据分段数量的影响，而Bitmap方法的计算速度则始终保持在高位，优势明显。统计结果表明，处理全年数据，Pycheron总耗时为59.3 d、ISPAQ为41.8 d，而Bitmap方法仅耗时1.27 h。这一结果进一步验证了Bitmap方法在复杂的真实数据场景下，进一步验证了Bitmap方法所具备的显著性能优势。

针对存在数据分段乱序、断记等异常情况的真实场景，以灾备中心2023年8月28日接收备份的AH.SHT.00.BHZ通道数据为例，进行计算结果对比分析。该数据存在明显的分段乱序和相互包含的现象，具体分段列表和异常说明列于表3。通过人工分析该日分段列表可知，主要异常包括：第3段起始时间晚于第4段，存在乱序；第5、7、8段与第3段完全重叠；全天仅在第1段前、后分别存在2.19 s和3.65 s的断记，占全天总时长的比例仅为0.0068%。该时段波形如图5所示。分别使用Pycheron、ISPAQ、Bitmap方法计算该日数据的连续性指标，结果显示：全天数据完整率分别是88.743 2%（Pycheron）、71.566 2%（ISPAQ）、99.993 2%（Bitmap）；断记次数分别是2、5、2。可见，Bitmap方法的计算结果与人工分析结论完全一致，而Pycheron和ISPAQ在完整率计算上均出现明显偏差，进一步验证了Bitmap方法在处理异常数据场景时具有更高的准确性和可靠性。

表3 AH.SHT.00.BHZ（2023年8月28日）数据分段列表

Table 3 Data segment list for AH.SHT.00.BHZ (August 28, 2023)

分段序号	分段起止时间	异常情况
1	2023-08-28T00:00:02.20—2023-08-28T19:51:26.34	
2	2023-08-28T19:51:30.00—2023-08-28T19:52:39.99	与第1段间存在断记
3	2023-08-28T19:53:32.00—2023-08-28T19:53:32.99	
4	2023-08-28T19:52:40.00—2023-08-28T22:35:34.99	乱序，包含第3段
5	2023-08-28T19:53:32.00—2023-08-28T19:53:32.99	与第3段完全重叠
6	2023-08-28T22:35:35.00—2023-08-29T00:00:00.00	
7	2023-08-28T19:53:32.00—2023-08-28T19:53:32.99	与第3段完全重叠
8	2023-08-28T19:53:32.00—2023-08-28T19:53:32.99	与第3段完全重叠

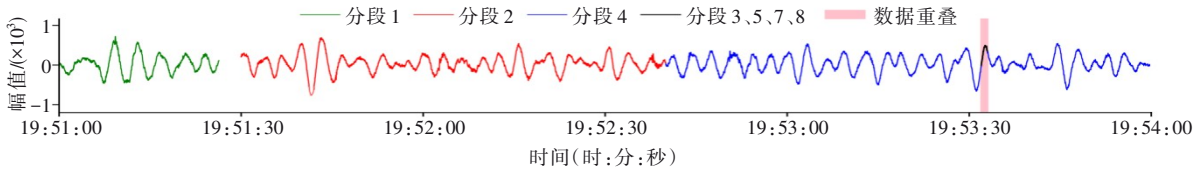


图5 AH.SHT.00.BHZ（2023年8月28日19:51:00—19:54:00）的波形

Fig. 5 Waveform of AH.SHT.00.BHZ (19:51:00—19:54:00, August 28, 2023)

灾备中心在2023年8月14日接收备份的JS.LIS.00.BHZ通道数据中，出现了更频繁的数据分段间相互包含、乱序以及大量头尾重叠的复杂情况。该日波形如图6所示，图中以不同颜色区别各数据分段。可以看出，当日数据分段数量众多，且存在许多小范围的重叠现象。表4给出了当日18:15至18:21的数据分段列表和异常说明。通过人工分析可知，在这短短6 min内出现了多次分段异常排列。例如，虽

然第3段与第4段之间存在10 s间隔，但第5段与第3段在时间上紧密衔接，实际上并未发生数据断记；当日其他时段出现的间隔情况也多与此类似，并非真正的断记。针对该6 min时段，Pycheron、ISPAQ和Bitmap方法计算出的断记次数分别为2次、3次和0次；而对全天数据进行计算，得到的断记次数则分别是42次、175次和5次。由此可见，Bitmap方法能够准确识别出实际并未发生断记的情况，而Pycheron和ISPAQ在断记次数的判定上存在明显偏差。此外，由于该日数据分段数多达859段，Pycheron和ISPAQ的处理耗时分别为3.69 s和6.40 s，而Bitmap方法仅耗时0.01 s，展示出显著的性能优势。

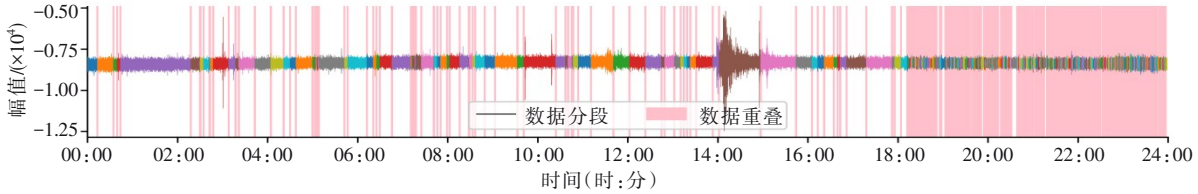


图6　JS.LIS.00.BHZ(2023年8月14日)的波形

Fig. 6　Waveform of JS.LIS.00.BHZ (August 14, 2023)

表4　JS.LIS.00.BHZ(2023年8月14日 18:15:00—18:21:00)数据分段列表

Table 4　Data segment list for JS.LIS.00.BHZ (18:15:00—18:21:00, August 14, 2023)

分段序号	分段起止时间	异常情况
1	2023-08-14T18:15:00.00—2023-08-14T18:15:20.99	
2	2023-08-14T18:15:11.00—2023-08-14T18:16:56.99	头部与第1段尾部重叠
3	2023-08-14T18:16:47.00—2023-08-14T18:17:38.99	头部与第2段尾部重叠
4	2023-08-14T18:17:49.00—2023-08-14T18:17:58.99	
5	2023-08-14T18:17:39.00—2023-08-14T18:18:38.99	乱序，包含第4段
6	2023-08-14T18:18:49.00—2023-08-14T18:18:58.99	头部与第5段尾部重叠
7	2023-08-14T18:18:39.00—2023-08-14T18:19:36.99	头部与第6段尾部重叠
8	2023-08-14T18:19:47.00—2023-08-14T18:19:49.99	
9	2023-08-14T18:19:37.00—2023-08-14T18:19:54.99	乱序，包含第8段
10	2023-08-14T18:19:50.00—2023-08-14T18:20:04.99	头部与第9段尾部重叠
11	2023-08-14T18:19:55.00—2023-08-14T18:21:00.00	头部与第10段尾部重叠

通过对数据集Ⅱ(中国大陆测震台网全年真实数据)的计算结果进行对比分析，以本文提出的Bitmap方法结果为基准，分别统计Pycheron和ISPAQ在完整率与断记次数两项关键指标上与基准结果的偏差。设定判定算法失效的阈值为：完整率差值超过5%或断记次数差值超过10次。统计结果显示，Pycheron在完整率计算上失效6 533次，在断记次数计算上失效15 237次；ISPAQ在完整率计算上失效6 039次，断记次数计算上失效25 507次。这进一步证实了现有方法在面对分段乱序、重叠等异常情况时存在鲁棒性不足的问题。相比之下，Bitmap方法从根本上避免了此类问题，其计算结果更为稳定可靠。

4　结论

本文针对现有地震波形数据连续性评价方法在应对数据分段数量过多、分段间存在乱序或重叠等异常排列情况时，所暴露出的计算效率低、结果可靠性不足等问题，提出了一种基于Bitmap的连续性评价方法。该方法将波形数据的每个采样点映射为表示记录状态和重叠状态的二进制位，通过构建Bitmap位向量，直观表征评价时间窗口内数据的连续状况，并基于此快速计算出数据完整率、断记与重叠次数及时长等核心指标。该方法从数据结构层面消除了异常分段对计算过程的干扰，从算法层面保障了结果的准确性与鲁棒性。通过在仿真数据和灾备中心备份的全国测震真实数据上进行对比实验，结果表明：与传统方法相比，本文所提方法在计算性能上具有显著优势，且对复杂数据场景(如大量分段、乱序、重叠等)表现出更强的适应性与结果一致性。目前，该方法已成功应用于中国大陆测震台网的日常数据质量评价业务，其核心思想亦可推广至强震动监测、地球物理观测等其他类型时间序列数据的连续性评价中，具有较好的应用前景。

参考文献(References)

[1] 许卫卫. 地震观测数据质量控制研究综述[J]. 地球物理学进展, 2018, 33(3): 998-1004.
XU Weiwei. Research progress on seismic data quality control [J]. Progress in Geophysics, 2018, 33(3): 998-1004.

[2] 孙业君. 地震台网记录数据质量评估的方法研究与实现[D]. 南京: 东南大学, 2015.
SUN Yejun. Research and implementation of quality evaluation method of seismic network recording data[D]. Nanjing: Southeast University, 2015.

[3] TRANI L, KOYMANS M, ATKINSON M, et al. WFCatalog: a catalogue for seismological waveform data[J]. Computers & Geosciences, 2017, 106: 101-108.

[4] PETERSEN G M, CESCA S, KRIEGEROWSKI M. Automated quality control for large seismic networks: implementation and application to the AlpArray seismic network[J]. Seismological Research Letters, 2019, 90(3): 1177-1190.

[5] BATINI C, CAPPIELLO C, FRANCALANCI C, et al. Methodologies for data quality assessment and improvement [J]. ACM Computing Surveys, 2009, 41(3): 1-52.

[6] MERITXELL G O, SIERRA B, FERREIRO S. On the evaluation, management and improvement of data quality in streaming time series[J]. IEEE Access, 2022, 10: 81458-81475.

[7] 陈通, 韩雪君, 马延路. 时序数据库在海量地震波形数据分布式存储与处理中的应用初探[J]. 中国地震, 2022, 38(4): 799-809.
CHEN Tong, HAN Xuejun, MA Yanlu. Preliminary application of time series database in distributed storage and processing of massive seismic waveform data[J]. Earthquake Research in China, 2022, 38(4): 799-809.

[8] PARK J, HAYWARD C, KIM B I, et al. Data quality control tools used to monitor seismoacoustic research arrays in South Korea[J]. Journal of Seismology, 2023, 27(4): 659-679.

[9] LEGENDRE C P, KUMAR U. Geohazards: analysis, modelling and forecasting[M]. Singapore: Springer Nature Singapore, 2023: 51-71.

[10] CASEY R, TEMPLETON M E, SHARER G, et al. Assuring the quality of IRIS data with MUSTANG[J]. Seismological Research Letters, 2018, 89(2A): 630-639.

[11] RINGLER A T, HAGERTY M T, HOLLAND J, et al. The data quality analyzer: a quality control program for seismic data[J]. Computers & Geosciences, 2015, 76: 96-111.

[12] AUR K A, BOBECK J, ALBERTI A, et al. Pycheron: a Python-based seismic waveform data quality control software package[J]. Seismological Research Letters, 2021, 92(5): 3165-3178.

[13] SHARER G, KEYSON L, TEMPLETON M E, et al. Developments in seismic data quality assessment using MUSTANG at the IRIS DMC [C]//AGU Fall Meeting Abstracts. New Orleans, LA: AGU, 2017: IN52A-04.

[14] ICHINOSE G. Waveform data quality assessment[R]. Livermore, CA, United States: Lawrence Livermore National Laboratory (LLNL), 2023.

[15] YULIATMOKO R S, GUNAWAN M T, ADI M, et al. Seismic station quality monitoring and evaluation system in Indonesia[J]. IOP Conference Series: Earth and Environmental Science, 2023, 1276(1): 012045.

[16] COLWELL J R, WITSIL A J C, MACPHERSON K A, et al. Unsupervised machine learning clustering of seismic and infrasound data quality metrics[J]. Seismological Research Letters, 2024, 95(3): 1812-1833.

[17] 郑秀芬, 欧阳飚, 张东宁, 等. "国家数字测震台网数据备份中心"技术系统建设及其对汶川大地震研究的数据支撑[J]. 地球物理学报, 2009, 52(5): 1412-1417.
ZHENG Xiufen, OUYANG Biao, ZHANG Dongning, et al. Technical system construction of Data Backup Centre for China Seismograph Network and the data support to researches on the Wenchuan earthquake[J]. Chinese Journal of Geophysics, 2009, 52(5): 1412-1417.

[18] CHAN C Y, IOANNIDIS Y E. Bitmap index design and evaluation[J]. ACM SIGMOD Record, 1998, 27(2): 355-366.

[19] 沈瑛, 陈望远, 侯晨煜, 等. 一种基于Bitmap的活动时间冲突查询算法[J]. 中南大学学报(自然科学版), 2018, 49(11): 2738-2744.
SHEN Ying, CHEN Wangyuan, HOU Chenyu, et al. A bitmap index based activities conflict query algorithm[J]. Journal of Central South University (Science and Technology), 2018, 49(11): 2738-2744.

[20] 刘阳, 张天石, 李世超, 等. 基于Bitmap的油水井采注优化实时推理引擎[J]. 自动化学报, 2017, 43(6): 1007-1016.
LIU Yang, ZHANG Tianshi, LI Shichao, et al. A real-time reasoning engine for injection-production optimization of water and oil wells on account of Bitmap[J]. Acta Automatica Sinica, 2017, 43(6): 1007-1016.

[21] 柴旭超, 王文青, 王丹宁, 等. 国家地震数据灾备中心大数据存储平台建设方案与研究[J]. 大地测量与地球动力学, 2017, 37(增刊3): 1-8.
CHAI Xuchao, WANG Wenqing, WANG Danning, et al. Big data storage platform construction scheme and research of the National Seismic Data Backup Center[J]. Journal of Geodesy and Geodynamics, 2017, 37(Suppl03): 1-8.

[22] KRISCHER L, MEGIES T, BARSCH R, et al. ObsPy: a bridge for seismology into the scientific Python ecosystem[J]. Computational Science & Discovery, 2015, 8(1): 014003.

(本文编辑:张向红)