

SEED 格式台站卷及带有可变头段节 数据记录的实例剖析

万永革 李鸿吉 姚立平 张德存

(中国北京 100081 国家地震局地球物理研究所)

摘要 地震数据交换标准 (SEED) 是国际地震数据交换的一种标准格式。本文简要介绍了台站磁带的 SEED 格式, 详细分析了带有可变头段节的数据记录的结构, 并给出了从带有可变头段节的数据记录中提取地震数据的实例。

主题词 地震数据; 数据处理

1 引言

地震数据交换标准 (Standard for the Exchange of Earthquake Data) 简称 SEED, 是国际地震数据交换的一种标准格式, 它可以帮助世界范围内的地震学家方便地使用和共享地震数据。中国数字地震台网除产生高质量的地震数据外, 还承担着国际地震数据交换的任务。

SEED 格式的逻辑卷分为现场台站卷、台站网络卷和事件网络卷。一般来说, 台站网络卷和事件网络卷是由现场台站卷编排和附加辅助描述信息而形成的。因此, 深入研究现场台站卷的记录格式无论是对于地震数据的实时处理还是对于世界范围的地震数据交换都具有一定的意义。

地震数据的原始存储介质为台站记录磁带, 是各台站的数据采集和处理系统的输出结果, 直接用于进行国际地震数据交换和地震数据处理。因此, 弄清了台站磁带的记录格式也就抓住了问题的关键。

王秀文等^[1]提供了地震数据交换标准的基本概念和各种逻辑卷结构, 对 SEED 卷的控制头段、时间序列 (数据记录) 和分区块给予了详细的说明。朱文林等^[2]对新一代 CDSN 台站磁带数据记录格式进行了深入剖析, 结合磁带内容详细阐述了台站卷的记录格式。我们在运用台站磁带提取地震数据的过程中, 对数据记录的连续性进行了检验并找到了朱文林等在实验中没有遇到的带有可变头段节的数据记录块。本文试图对记录块的时间连续性进

国家地震局地球物理研究所论著编号 95A0100.

本文收到日期: 1995-05-09.

行说明, 并对带有可变头段节的数据记录格式进行实例剖析。

2 台站磁带记录格式

台站磁带在 SEED 格式中称为台站网络卷。CDSN 台站卷仅存时间序列数据。台站磁带组成的基本单元为数据记录(record)。一个数据记录由一个固定头段节(含数据记录标识块的个数)、可变头段节和数据节组成。

固定头段节(长度为 64 字节)

可变头段节(长度不固定)

第一压缩包(长度为 192 字节)

第二压缩包(长度为 192 字节)

⋮

数据记录的固定头段节长度为 64 个字节, 其中前 48 个字节均有固定的意义。固定头段节的前 20 个字节使用二进制码。固定头段节对本数据记录的台网、台站和通道给出标识并说明记录开始时间、采样、数据质量等。对此文献 [1]、[2] 均有详细阐述。

数据记录的可变头段节位于固定头段节之后、数据节之前。是由若干可选分区块组成的, 它们提供了各种事件检测、标定手段等信息, 用以描述数据的性质。可变头段节使用的是二进制码。

数据记录的数据节为时间序列数据, 位于可变头段节之后。数据节是由若干压缩包组成, 每个压缩包的长度为 192 个字节。压缩包的结构为: 控制信息描述(4 字节), 前向积分常数(4 字节), 后向积分常数(4 字节), 地震数据差值(52 字节)(以上 64 字节为第一个压缩框); 控制信息描述(4 字节), 地震数据差值(60 字节)(以上 64 字节为第二个压缩框); 控制信息描述(4 字节), 地震数据差值(60 字节)(以上 64 字节为第三个压缩框)。

地震信号具有高度相关性, 时间样本序列的采样差值非常小。在 99% 的时间内, 地震信号的样本差值可用一个字节来表示。SEED 格式巧妙地利用了地震信号的这个特点进行数据压缩。它们采用 Steim 压缩算法, 其原理如下。

设某压缩包的原始样本序列为 $X_0, X_1, X_2, \dots, X_n$, 据差值公式

$$D_n = X_n - X_{n-1}$$

则

$$D_0 = X_0 - X_1, D_1 = X_1 - X_2, \dots, D_n = X_n - X_{n-1}$$

这里, X_0 为前向积分常数; X_n 为后向积分常数。分别和上面介绍的压缩包结构相对应。

对于不能用一个字节表示的差值, 可用 2 个字节或 4 个字节表示。由于这种差值位置的不确定性, 不可能在头段对其进行说明。为此, SEED 格式记录的压缩包数据每隔 64 字节有一控制描述信息, 一个 64 字节称为一个数据框(Frame), 一个压缩包中有 192 个字节, 共 3 个压缩框。每个数据框的头 4 个字节为控制描述信息, 描述本数据框表示差值的字节个数。这 4 个字节(byte) 总共 32 位(bit), 每两位控制 4 个字节, 共控制 64 个字节。

需要指出的是, 台站磁带 SEED 格式的数据记录大部分为不带可变头段节的数据记

录，朱文林等^[2]剖析的数据记录实例即为不带可变头段节的数据记录。

3 数据记录各通道存储的顺序

我们任意抽出北京台网记录的一盘磁带，提取了数据头段的某些标定信息特别是通道信息和时间信息。由于台站在地震数据记录过程中，将各频段的各通道地震数据分别分配一缓冲区，一旦地震数据长度占满一个数据记录（连同数据头段达 4 096 字节），就写进磁带。各通道的采样率及其相互关系就决定了其数据记录块的多少和顺序（表 1）。

表 1 台站磁带数据记录的部分头段信息

序号	信道	记录开始时间	样本数	采样率
		年-日-时-分-秒-万分之一秒		
000001	SHE	1994-236-5-46-25-8690	1832	40
000002	SHZ	1994-236-5-46-40-3440	1830	40
000003	SHN	1994-236-5-47-08-3190	1822	40
000004	SHE	1994-236-5-47-11-6690	1818	40
000005	BHE	1994-236-5-44-58-7090	3612	20
000006	BHN	1994-236-5-45-02-3090	3612	20
000007	SHZ	1994-236-5-47-26-0940	1820	40
000008	SHN	1994-236-5-47-53-8690	1824	40
000009	SHE	1994-236-5-47-57-1190	1822	40
000010	SHZ	1994-236-5-48-11-5940	1822	40
000011	LHZ	1994-236-5-16-43-5590	1830	1
000012	SHN	1994-236-5-48-39-4690	1822	40
000013	SHE	1994-236-5-48-42-6690	1816	40
000014	BHZ	1994-236-5-46-32-9090	3612	20
000015	SHZ	1994-236-5-48-57-1440	1822	40
000016	SHN	1994-236-5-49-25-0190	1818	40
000017	SHE	1994-236-5-49-28-0690	1834	40
000018	SHZ	1994-236-5-49-42-6940	1836	40
000019	SHN	1994-236-5-50-10-4690	1844	40
000020	SHE	1994-236-5-50-13-9190	1814	40

4 从带有可变头段节的数据记录中得到数据

我们从磁带中找到了一块具有代表性的带有可变头段节的数据记录，此数据记录是北京台记录的 1994 年 365 天 2 点的北部湾地震记录的一小部分。

(1) 由于磁带数据记录除头段的 20 个字节为可见的 ASCII 外，其余全为不可见的十六进制码，要搞清数据记录的每一个细节，首先要把不可见的十六进制编码变为可见的十六进制数，为此，我们编写了用 C 语言直接对磁带的数据记录进行转换的程序 rdhex.c，其抽取的实例转换结果见表 2。

表 2 rdhex.out 对实例数据的转换结果及说明

转 换 的 数 据				说 明	
30333137	39304420	424a5420	20202042	固定头段节 共 64 字节	
484e2020	07ca016d	023b1400	0ae60d0e		
00140001	40000001	00000000	01000040		
fffffd13	05552200	00022900	00000000		
00c90000	46156c00	3fcccccd	43a48000	可变头段节 共192字节	
010007ca	016d0302	010014aa	01050909		
09000101	fff6ede2	fc060f11	eae2fc01		
fcf2edf8	f5f700fe	f7020df3	f003f5f1		
15555555	0a120419	0de7fafd	f2eff91d		
100806ee	02fbed06	0ff5elec	001a190a		
05e9e7f5	07170b04	f5f0fafb	feff0f0d		
f1ecf0fd	13fef6fc	010af9f7	0508fff0		
15555555	edf71d1b	f2afff11	f510280f		
eaef070a	0f18faf9	f7fe1811	1f180909		
f4f9f4ff	161e2411	05f6ecff	0f0c0b0b		
fffe01ed	f31010fc	f3f9f202	1602f6e8		
01555555(控制)	00000033(前向)	000000a5(后向)	0bf90b19	第一数据框 64 字节	第一 压缩包，共 192 字节
01040a12	05f3fa05	05fe0615	1502f5f1		
fa0f1d1c	fff6ede2	fc060f11	eae2fc01		
fcf2edf8	f5f700fe	f7020df3	f003f5f1		
15555555(控制)	0a120419	0de7fafd	f2eff91d		
100806ee	02fbed06	0ff5elec	001a190a		
05e9e7f5	07170b04	f5f0fafb	feff0f0d		
f1ecf0fd	13fef6fc	010af9f7	0508fff0		
15555555(控制)	edf71d1b	f2afff11	f510280f		
eaef070a	0f18faf9	f7fe1811	1f180909		
f4f9f4ff	161e2411	05f6ecff	0f0c0b0b		
fffe01ed	f31010fc	f3f9f202	1602f6e8		
01555555(控制)	0000008a(前向)	0000003f(后向)	e50708fa	第一数据框 64 字节	第二 压缩包，共 192 字节
0f11e4d8	f6f7ff17	08eae01	fef7f2fd		
04fff9e4	f006f2f9	160efbe9	e0edfc08		
0f0af9f2	dfdb0e1a	1008ebf1	ecf11812		
15555555(控制)	081719ea	d5f4ec02	3222f8f6		
1509f403	0d13201c	07faf403	08ff191c		
0b05f906	0a030611	0bf7f702	07ff0820		
09fafde9	ff1b08f1	0825f8da	0106f0f5		
15555555(控制)	f9f5f809	09e9e7eb	f20b00f9		
f9f1eeec7	fc0709fd	e3e1e3ea	03150e05		
edd2f01b	1f1813f2	e5efe6f8	10261f00		
06fcf2f0	ef151c1f	0ff6fff7	0210061a		
...					

一个数据记录 (record) 长度为 4 096 字节。本例中, 前 64 字节为可变头段节, 紧接

着是占有一个压缩包长度(192字节)的可变头段节,其余的为数据节,长度为3840字节,分为20个压缩包存储。表2还对压缩包的结构结合实例进行了说明,其中,“控制”代表控制描述信息,“前向”代表前向积分常数,“后向”代表后向积分常数。

由表2可见,可变头段节与数据压缩包长度是一样的,均占192字节。它们的主要区别在于控制描述信息位置处的数据。压缩包中的控制信息占4个字节,相对于第一个压缩包的第一个压缩框,控制信息的数据为01555555,共32bit,可分解为16个两bit,分解为00,00,00,01,01,01,01,01,01,01,01,01,01,01,01,01,每两个bit控制4个字节,在SEED格式中规定:00表示所控制的4个字节为非数据项或特殊信息,01表示所控制的4个字节为4个1字节表示的差值,10表示所控制的4个字节为2个2字节表示的差值,11表示所控制的4个字节为1个4字节表示的差值。结合本例,第1个00说明本数据框的1—4字节为特殊信息,即控制描述信息;第2个00说明5—8字节为4字节表示的前向积分常数;第3个00说明9—12字节为4字节表示的后向积分常数;第1个01按规定表示13—16字节为4个1字节差值;以此类推。

若将可变头段节看成第一个压缩包,则第一组控制描述信息所在位置的数据为00c90000,分解为00,00,11,10,10,01,00,00,00,00,00,00,00,00,00,00,按照SEED格式中的规定,不能有这样的控制描述信息。因此,不能将可变头段节按数据节处理。

(2)SEED格式中的头段包含着丰富的说明信息,提取所有的固定头段信息对理解本数据记录的结构是很有帮助的。朱文林等^[2]已详细说明了固定头段节的每个字节所表示的内容并结合实例给出了说明,在此从略。我们编写了提取固定头段全部信息的程序getheader.c,所得上例的固定头段信息如下:

The Order number of record is: 031790

The Header Flags is: D

The station name is: BJT

The location is:

The channel is: BHN

The network is:

The year is 1994

The day is 365

The hour is 2

The minite is 59

The second is 20

The fraction of second is 2790

The number of sample is 3342

The factor of sample rate is 20

The sample rate multiplier of record is 1

The activity flags record is 64

The I/O flags of record is 0

The quality flags of record is 0

The Number of all continue blockette continuing of record is 1

The value of time correction of record is 0

The number of begining bytes is 256

The first blockette of record is 64

需要特别指出的与一般数据记录不同的是后续标识块总数为 1 而不为 0, 数据开始字节编号为 256 而不为 64, 正好多了 192 个字节相当于一个压缩包的长度。

(3) 为了检验我们的认识, 我们编写了提取原始数据的程序 `getdata.c` 并对本数据记录进行操作, 所得结果与在 DMC (CDSN 数据管理中心) 利用实用程序 `getdata` 得到的同一段数据完全吻合。

需要特别指出的是每一个压缩包的第一个差值 (即后向积分常数后字节表示的第一个差值) 为本压缩包的第一个真实地震数据 (前向积分常数所表示的数据) 与上一个压缩包的最后一个真实地震数据的差, 对于第一个压缩包, 则是本数据记录的第一个地震数据与上一个数据记录的最后一个地震数据的差。这样不仅压缩包之间不间断 (没有 gap 出现), 数据记录之间的数据也连接起来了 (参见第 2 节台站磁带记录格式)。

5 结 论

对以上讨论分析, 我们可得出以下结论。

(1) 台站磁带数据的基本存储单元为第 2 节所示的数据记录 (record), 数据记录的长度为 4 096 字节。

(2) 由第 2 节的顺序排列的数据记录头段信息可知, 磁带数据记录各频段混合存储, 即每个频段的通道数据记录够一个数据记录 (record) 就写进磁带。

(3) 每个数据记录分为 3 个区。第 1 个区为固定头段节区, 长度为 64 字节。第 2 个区为可变头段节区, 长度不固定。第 3 个区为数据存储区。

(4) 数据存储区的组成单元为压缩数据包, 每个压缩包以 Steim 压缩算法存储。每个压缩包又分为 3 个压缩框, 每个压缩框含有一控制描述信息, 对本压缩包的所有字节的差值分配给予描述。

本项工作得到了 CDSN 数据中心多位同志的大力支持和协助, 特别是朱文林同志给本项工作提供了很多方法和建议, 在此作者谨致谢意。

参考文献

- [1] 王秀文、姚立平、赖德伦、牟其铎, 1994. 地震数据交换标准, 地震地磁观测与研究, 15, 2.
- [2] 朱文林、张 昱、曹永键、牟其铎, 1994. 新一代 CDSN 台站磁带数据记录格式, 地震地磁观测与研究, 15, 3.
- [3] 姚立平, 1993. 地震数据存储交换中的压缩技术, 地震地磁观测与研究, 14, 5.

The SEED format of station tape and the instance revealment of the data record with variable header section

Wan Yongge, Li Hongji, Yao Liping and Zhang Decun

(Institute of Geophysics, State Seismological Bureau, Beijing 100081, China)

Abstract

SEED is the abbreviation of Standard for Exchange of Earthquake Data. In this paper, we give a brief description of the SEED format of station tape, analyse the structure of the data record with variable header section in detail, and give a example of retrieving data from the record with variable header section.

Key words: seismic data, data processing