

基于多分辨率内插预测的声波勘探数据无损压缩¹

吴乐南 王 桥

(东南大学无线电工程系 南京 210096)

摘 要 提出一种基于多分辨率内插预测 (MRIP) 的数据压缩新方法, 可将海洋地球物理勘探数据和声波测井数据无损压缩 50% 以上。

关键词 数据压缩, 无失真编码, 内插预测, 勘探信号

中图分类号 TN911.2

1 引言

石油地球物理勘探数据浩如烟海, 压缩存储意义重大, 但通常只有完全无失真的压缩方法才更能为用户所接受。问题是, 此类数据很难无损压缩, 例如对于海洋地球物理勘探数据^[1], 商用 ARJ 软件只能压缩掉 3.4%, 几乎失效。文献 [1] 采用按系数不变法设计的线性预测器与霍夫曼编码, 达到了 1.67 倍的平均无失真数据压缩比 (CR), 而且简单快速, 可以在线处理。本文则提出一种基于多分辨率内插预测 (MRIP) 的数据压缩新方法, 以牺牲处理速度为代价, 达到了平均 $CR > 2:1$ 的无失真压缩效果, 可用来减半巨大的数据存储费用。

2 内插预测的导出

通常的差值脉冲编码调制 (DPCM) 方法是基于某种顺序预测模型^[2], 它只利用了过去的因果性采样值来预测当前的数据样本, 未能考虑当前点与后续数据之间的相关性。而从事后处理的角度看, 完全有可能利用尽可能多的数据 (因果的和非因果的) 以得到更好的预测效果 (从而得到更好的压缩结果)。

我们设想把一条勘探记录的 N 个采样数据 $\{x(n), n = 1, 2, \dots, N\}$ 按序号的奇偶一分为二, 丢掉 $N/2$ 个偶数点 $\{x(2k), k = 1, 2, \dots, N/2\}$, 然后设法用 $N/2$ 个奇数 $\{x(2k-1) = y(k), k = 1, \dots, N/2\}$ 来预测它们。显然可利用相邻奇数点内插得到偶数点, 但这种简单插值的效果不会太好。而对于带限 (低通) 信号的采样, 有精确的内插恢复公式^[2]:

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \frac{\sin[(t - nT_s)\pi/T_s]}{(t - nT_s)\pi/T_s}, \quad (1)$$

其中 T_s 为采样间隔。(1) 式表明可通过内插函数把离散信号 $x(nT_s)$ (简记为 $x(n)$) 恢复为连续信号 $x(t)$ 。对 $y(k)$ 利用 (1) 式并取 $t = (k + 0.5)T_s$ 为 $y(k)$ 的插值点, 则有

$$y(k + 0.5) = \sum_{n=-\infty}^{\infty} y(n) \frac{\sin[(k + 0.5 - n)\pi]}{(k + 0.5 - n)\pi} = \sum_{n=1}^{N/2} y(n) \frac{(-1)^{k-n}}{2(k-n)+1}, \quad k = 1, 2, \dots, N/2. \quad (2)$$

利用 (2) 式即可由 $N/2$ 个 $y(k) = x(2k-1)$ 值内插出 $N/2$ 个 $y(k + 0.5)$ 即 $x(2k)$ 的估计值 $x'(2k)$ 。由于 $y(k)$ 的采样率只有 $x(n)$ 的一半, 其频谱有可能产生混叠, 使得 $x'(2k) \neq x(2k)$, 因而我们将 $x'(2k)$ 看成 $x(2k)$ 的“预测值”, 对“预测误差”:

¹ 1998-08-26 收到, 1999-05-09 定稿
国家自然科学基金 (69772025) 资助课题

$$e(k) = x(2k) - x'(2k) \quad (3)$$

再进行熵编码, 从而实现无失真压缩。这就是我们所谓的内插预测。

接收端将解码后的预测误差 $e(k)$ 与自身同样根据 (2) 式“预测”出的 $x'(2k)$ 相加, 即可恢复偶数点数据 $\{x(2k), k = 1, 2, \dots, N/2\}$ 。

3 多分辨率内插预测

采用与文献 [1,2] 相同的声波勘探数据, 由 (2)、(3) 式得到 $N/2$ 个偶数点数据的内插预测误差信号的熵值如表 1, 可见效果相当显著, 这启示我们尝试继续对该 $N/2$ 个奇数点数据再次进行 $N/4$ 点的内插预测; 对 $N/4$ 个奇数点再次进行 $N/8$ 点的内插预测; \dots ; 由此便得到了由底向上数据分辨率递减的所谓多分辨率的内插预测编码方法。可以预见, 随着分辨率的降低, 数据样本间的相关性越来越小, 预测结果也会越来越差, 直至得不偿失。由表 1 可见, 当分解到 $N/8$ 层次 (对于测井信号分解到 $N/4$ 层次) 便不再有效果了, 应该到此为止了。也就是说, 对于声波勘探数据, 只需进行 2 ~ 3 级内插预测即可。

表 1 两类勘探信号的多分辨率内插预测结果

单位: bit/ 样本	回波信号 1	回波信号 2	测井信号
N 点原始数据的熵值	10.994	11.004	6.570
$N/2$ 个偶数点误差信号的熵值	5.224	3.792	3.028
$N/2$ 个奇数点原始数据的熵值	10.123	10.161	6.146
奇、偶平均后的 N 个数据的熵值	7.674	6.977	4.587
$N/4$ 个偶数点误差信号的熵值	5.978	5.755	5.888
$N/4$ 个奇数点原始数据的熵值	9.192	9.221	5.706
奇偶平均后的 $N/2$ 个数据的熵值	7.585	7.488	5.797
与第一级 $N/2$ 点平均后的总熵值	6.405	5.640	4.413
$N/8$ 个偶数点误差信号的熵值	8.266	8.266	5.792
$N/8$ 个奇数点原始数据的熵值	8.258	8.256	5.115
奇偶平均后的 $N/4$ 个数据的熵值	8.262	8.261	5.454
与第一、二级平均后的总熵值	6.172	5.400	4.350

4 编码及压缩结果

表 1 中各栏的最后一行是对该级分解内插预测误差值及剩余数据进行熵编码所能达到的无失真压缩码率的下限, 可见, 如果采用的熵编码效率很高, 则只需采用一次内插预测即可得到 2:1 的数据压缩比。为简单起见, 我们对海洋地球物理勘探数据和声波测井数据均按 $N/2$ 、 $N/4$ 和 $N/8$ 偶数点内插预测误差的统计特性分别设计了简化的霍夫曼码表 (因为预测误差的动态范围即所需要表示的可能的事件数远大于可用于统计的数据样本数), 但实测发现霍夫曼码对于第三级改进不大, 故实际只取两级内插预测, 得到的压缩结果如表 2。

表 2 两类勘探信号的 MRIP 实际编码压缩结果

单位: bit/ 样本	回波信号 1	回波信号 2	测井信号
$N/2$ 个偶数点误差信号的编码率	5.737	4.064	3.310
$N/4$ 个偶数点误差信号的编码率	6.372	6.115	7.200
$N/4$ 个奇数点原始数据的编码率	14.718	14.718	7.987
奇偶平均后 $N/2$ 个奇数点编码率	10.545	10.417	7.594
N 个数据奇偶平均后的编码率	8.141	7.241	5.447

与表 1 的理论值相比, 表 2 结果并不理想, 特别是对于 $N/4$ 个奇数点原始数据的编码效率太低, 还有改进的潜力。但即便如此, 对于海洋地球物理勘探数据的平均无失真压缩比也达到了 $16/(8.141+7.241)/2=2.08$; 对于声波测井数据则达到了 $12/5.447=2.20$ 。

5 结 论

本文提出并实际采用了两级分辨率内插预测加简化霍夫曼编码的数据压缩新方法, 使得对海洋地球物理勘探数据和声波测井数据的平均无损压缩比突破了 2:1, 代价是增加了预测器的阶数并需要用浮点表示。但由于可完全无损地压缩一半数据, 而且计算机的处理速度越来越高, 因此用“时间”换“空间”越来越可行。另外, 该预测算法的结构本身也适合于用硬件并行或流水处理。对熵编码的改进 (例如采用算术编码) 还可望进一步提高数据压缩比。

参 考 文 献

- [1] 吴乐南. 声波勘探信号的预测编码. 石油地球物理勘探, 1995, 30(4): 505-508.
- [2] 吴乐南. 数据压缩的原理与应用. 北京: 电子工业出版社, 1995, 343-348.

LOSSLESS CODING OF ACOUSTIC EXPLORATION WAVES BY MULTI-RESOLUTION INTERPOLATION AND PREDICTION

Wu Lenan Wang Qiao

(*Department of Radio Engineering, Southeast University, Nanjing 210096*)

Abstract A novel encoder based on interpolation and prediction in multi-resolution levels is proposed, it can compress ocean geophysical exploration echoes and sound-logging data losslessly more than 50 per cent.

Key words Data compression, Lossless coding, Interpolated predictions, Exploration signals

吴乐南: 男, 1952 年生, 教授, 博士生导师, 系副主任, 现从事数据压缩、视频处理及多媒体通信方面的教学与科研工作。

王 桥: 男, 1966 年生, 副教授, 现从事微局部分析、调和分析及其在信号处理中的应用。