

LZ77与LZ78

维基百科，自由的百科全书

LZ77与**LZ78**是Abraham Lempel与Jacob Ziv在1977年以及1978年发表的论文中的两个无损数据压缩算法。这两个算法是大多数LZ算法变体如LZW、LZSS以及其它一些压缩算法的基础。与最小冗余编码器或者行程长度编码器不同，这两个都是基于字典的编码器。LZ77是“滑动窗”压缩算法，这个算法后来被证明等同于LZ78中首次出现的显式字典编码技术。

目录

LZ77

LZ78

参考文献

外部链接

LZ77

LZ77算法通过使用编码器或者解码器中已经出现过的相应匹配数据信息替换当前数据从而实现压缩功能。这个匹配信息使用称为*长度-距离对*的一对数据进行编码，它等同于“每个给定长度个字符都等于后面特定距离字符位置上的未压缩数据流。”（“距离”有时也称作“偏移”。）

编码器和解码器都必须保存一定数量的最近的数据，如最近2 KB、4 KB或者32 KB的数据。保存这些数据的结构叫作*滑动窗口*，因为这样所以LZ77有时也称作滑动窗口压缩。编码器需要保存这个数据查找匹配数据，解码器保存这个数据解释编码器所指代的匹配数据。所以编码器可以使用一个比解码器更小的滑动窗口，但是反过来却不行。

许多关于LZ77算法的文档都将长度距离对描述为从滑动窗“复制”数据的命令：“在缓冲区中回退距离个字符然后从那点开始复制长度个字符。”尽管对于习惯于指令式编程的人们来说很直观，但是它仍然使得人们很难理解LZ77编码的一个特点：也就是说，长度距离对中的长度超过距离这样一种情况不仅是可接受的而且还是经常出现的情况。作为一个复制命令，就会让人费解：“回退一个字符然后从那点开始复制七个字符。”但是如果缓冲区中只有一个字符的话那该如何复制七个字符呢？然而将长度距离对看作对于特性的描述就可以避免这种混淆：后面的七个字符与前面的七个完全相同。这就意味着每个字符都可以通过在缓冲区找到确定下来：即使当前数据对解码开始的时候所要查找的字符并不在缓冲区中也可以。通过这个定义，这样的数据对将是距离字符序列的多次重复，也就是说LZ77成了一个灵活容易的行程长度编码形式。

尽管所有的LZ77算法都是根据同样的基本原理工作，但是如何输出编码后的数据可能会大不一样，例如长度与距离的取值范围是多少以及如何区分长度距离对和*字符面符号*（即直接用字符本身，而不是用长度距离对表示）。下面是一些实例：

- 在Lempel与Ziv最初1977年论文中描述的算法每次将数据输出成三个数值：在缓冲区中找到的最大匹配长度与位置以及紧随这个匹配的字符面符号。如果输入流中的两个相邻字符只能编码成字符面符号的话，那么长度就是0。
- 在PalmDoc格式中，长度距离对经常用两字节序列进行编码，在这两字节的16位中，11位表示距离，3位表示长度，剩余的两位用来表示第一个字节是这样一个两个字节序列的开头。

- 直到2004年，最流行的基于LZ77的压缩方法是同时使用了LZ77与哈夫曼编码的DEFLATE。字面符号、长度以及用于指示当前数据块结束的符号都放到一个字母表中。距离可以安全地放到一个单独的字母表中，由于距离只会出现在一个长度后面，所以它不可能与其它类型的符号混淆。

LZ78

LZ77算法针对过去的数据进行处理，而LZ78算法却是针对后来的数据进行处理。LZ78通过对输入缓存数据进行预先扫描与它维护的字典中的数据进行匹配来实现这个功能，在找到字典中不能匹配的数据之前它扫描进所有的数据，这时它将输出数据在字典中的位置、匹配的长度以及找不到匹配的数据，并且将结果数据添加到字典中。

尽管在最初得到流行，但是后来LZ78的普及逐渐衰减，这可能是由于在刚LZ78出现的一些年份，一部分LZ78算法获得了美国专利保护。最流行的LZ78压缩形式是LZW算法，这个算法是Terry Welch所开发的一个LZ78变体。

参考文献

- Jacob Ziv and Abraham Lempel; *A Universal Algorithm for Sequential Data Compression* (http://www.cs.duke.edu/courses/spring03/cps296.5/papers/ziv_lempe1_1977_universal_algorithm.pdf), IEEE Transactions on Information Theory, May 1977.

外部链接

- LZ77算法及变体的库、论文及源代码列表 (<http://www.compression-links.info/LZSS>)

取自 “<https://zh.wikipedia.org/w/index.php?title=LZ77与LZ78&oldid=66317555>”

本页面最后修订于2021年6月28日 (星期一) 10:47。

本站的全部文字在知识共享 署名-相同方式共享 3.0协议之条款下提供，附加条款亦可能应用。（请参阅使用条款）
Wikipedia®和维基百科标志是维基媒体基金会的注册商标；维基™是维基媒体基金会的商标。
维基媒体基金会是按美国国内税收法501(c)(3)登记的非营利慈善机构。