任意概率分布下 Golomb 码和扩展 Gamma 码的性能分析

杨胜天　　仇佩亮
（浙江大学信息与电子工程学系　杭州　310027）

摘　要：以信源的平均值给出了任意概率分布下 Golomb 码的平均码长的上下界和最优的参数选择准则。在 Golomb 码的基础上，进一步推广了 Elias 的 γ 码，提出了扩展的 γ 码，同时给出了其性能界和最优的参数选择准则。扩展 γ 码是一类通用码，而且在一定的条件下可以达到渐近最优的性能。最后，提出了一个低复杂性的基于 Golomb 码和扩展 γ 码的通用数据压缩框架，并通过构建一个样例系统说明了该数据压缩框架的实际应用价值。

关键词：信源编码，整数编码，Golomb 码，Elias γ 码

中图分类号：TN911.21　　**文献标识码**：A　　**文章编号**：1009-5896(2005)04-0514-05

Performance Analysis of Golomb Codes and Extended Gamma Codes for Arbitrary Probability Distributions

Yang Sheng-tian　　Qiu Pei-liang
(Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, *China)*

Abstract　The upper and lower bounds of the average codeword length of Golomb codes for arbitrary probability distributions as well as an optimal rule for choosing parameters are given in terms of the mean of sources. Furthermore, a class of extended gamma codes which are the generalization of Elias gamma code is constructed based on Golomb codes. The performance bounds and an optimal rule for choosing parameters are also given. Extended gamma codes are universal and can achieve asymptotically optimal performance under some conditions. Finally, a low complexity universal data compression framework based on Golomb codes and extended gamma codes is presented, and a sample system is constructed to indicate the significance of the data compression framework in practice.

Key words　Source coding, Integer coding, Golomb codes, Elias gamma code

1　引言

Golomb 码是 Golomb 于 1966 年提出的一类带参数的前缀码[1]，由一个一元码和一个匹配均匀分布的 Huffman 码构成。由于其编码和解码的算法十分简单，到目前为止，已被广泛地应用于数据压缩领域，如无损图像压缩[2, 3]。关于 Golomb 码的性能，Gallager 等证明了其在单边几何分布（One-Sided Geometric Distribution, OSGD）下是最优的[4]，并给出了参数为θ 的 OSGD 下的最优参数准则，即满足$\theta^{l}+\theta^{l+1}\le 1<\theta^{l}+\theta^{l-1}$ 的整数 l；Szpankowski 则进一步分析了其在 OSGD 下码字冗余长度的渐近特性[5]。另一方面，Merhav 等在双边几何分布（Two-Sided Geometric Distribution, TSGD）下提出了一类性能最优的扩展的 Golomb 码[6]。与这些理论上的进展相比，在实际中，Golomb 码的应用并没有仅仅局限于 OSGD 或 TSGD 的信源，它常常被用于那些具有近似递减概率特性但确切概率模型未知的信源上[3, 7, 8]，因此，很有必要对 Golomb 码在任意分布下的性能进行分析。

本文将在第 2 节给出 Golomb 码在具有有限均值的任意分布下的平均码字长度的上下界，并给出 Golomb 码达到最低上下界的参数公式；在第 3 节中，我们将在 Golomb 码的基础上构造一类扩展的 γ 码，并分析了这类码的性能界，进而证明这类码在非递增概率分布下为通用码，而且可以依据信源的有关统计信息选择合适的参数，从而获得渐近最优的性能；在第 4 节中，我们以 Golomb 码和扩展 γ 码为基础，提出了一个低复杂性的基于整数编码的通用数据压缩框架，并给出了一个简单的数据压缩系统构建实例和实验结果，进一步说明了 Golomb 码、扩展 γ 码以及该数据压缩框架的实际应用价值。以下是全文中使用的符号与定义。

首先，定义集合 $B=\{0,1\}$，并定义 B^*表示由集合 B 中的符号组成的所有有限长度的二进制序列的集合，集合 B^* 中的任意两个序列α和β的级联记为$\alpha\beta$，序列α的长度记为$|\alpha|$。设一组由可数无穷个码字构成的前缀码 $C=\{c_i\in B^*: i\in$

2003-09-18 收到，2004-05-20 改回

$N^+\}$，其中 $N^+=\{1,2,\cdots\}$。

其次，定义信源 $S=(N^+,P)$，N^+为信源的字符集，P 为定义在正整数集合上的概率分布函数。于是，信源 S 的熵和均值分别表示为 $H(P)\equiv E_P[-\log p(i)]$和$\mu_P\equiv E_P[i]$。（本文中，$\log x\equiv\log_2 x$，$\lfloor x\rfloor$表示对 x 下取整，$\lceil x\rceil$表示对 x 上取整，$\langle x\rangle\equiv x-\lfloor x\rfloor$，$x \bmod y\equiv x-\lfloor x/y\rfloor y$。）

我们将每个码字 c_i 与每个正整数联系起来，那么对于一个给定的前缀码 C 和一个给定的信源 S，其平均码长为

$$\overline{L_P(C)}\equiv E_P[|c_i|]=\sum_{i\in N^+}p(i)|c_i| \tag{1}$$

参照文献[9]的定义，我们称信源 S 的概率为非递增的，是指信源 S 的概率满足

$$p(1)\ge p(2)\ge\cdots\ge p(i)\ge\cdots \tag{2}$$

而将满足式(2)且具有正的有限信源熵的所有信源S定义为信源类 $\mathcal{S}_B$。前缀码 C 是通用的，当且仅当比率：

$$\frac{\overline{L_P(C)}}{\max\{1,H(P)\}}\le K_C\,,\quad \forall S\in\mathcal{S}_B \tag{3}$$

前缀码 C 是渐进最优的，当且仅当比率：

$$\frac{\overline{L_P(C)}}{\max\{1,H(P)\}}\le R_C(H(P))\le K_C,\text{且}\lim_{H(P)\to\infty}R_C(H(P))=1 \tag{4}$$

最后，我们还需要定义一些表示二进制序列的记号。令$(j)_2$为一个正整数j 的标准二进制表示（最左边的比特位必为 1）。例如，$(11)_2=1011$。又令$[j]_2$表示正整数j 的去除最高位的标准二进制表示。例如，$[11]_2=011$。$(j)_2$和$[j]_2$的长度分别记为

$$\Lambda(j)\equiv|(j)_2|=\lfloor\log j\rfloor+1\,,\quad \lambda(j)\equiv|[j]_2|=\lfloor\log j\rfloor \tag{5}$$

令$(j)_1$为一个正整数j 的一元表示。例如，$(1)_1=0$，$(2)_1=10$，$(3)_1=110,\cdots$。显然，$|(j)_1|=j$。令 $B_l(j)$为表示 0 到 l-1 这 l 个整数的分组码，其编码规则为

$$B_l(j)\equiv\begin{cases}B_{2^{k-1}}(j), & j<2^k-l\\ B_{2^k}(j+2^k-l), & j\ge 2^k-l\end{cases} \tag{6}$$

式中 $B_{2^k}(j)$ 定义为整数j 的 k 比特二进制表示，$k=\lceil\log l\rceil$。码字 $B_l(j)$ 的长度为

$$|B_l(j)|=\lceil\log l\rceil-u(2^{\lceil\log l\rceil}-l-1-j) \tag{7}$$

式中 $u(x)$为阶跃函数（当 $x\ge 0$，$u(x)=1$，否则 $u(x)=0$）。

2　任意分布下 Golomb 码的性能分析

本节将给出 Golomb 码在具有有限均值的任意分布下的平均码字长度的上下界，并给出使 Golomb 码达到最低上下界的参数公式。原始的 Golomb 码是基于非负整数的，本文为与描述整数码的文献保持一致，所以给出了以下这个略微修改的基于正整数的 Golomb 码。

定义参数为 l（$l\in N^+$）的 Golomb 码为

$$G_l=\left\{g_l(i):g_l(i)=\left(\left\lfloor\frac{i-1}{l}\right\rfloor+1\right)_1 B_l((i-1)\bmod l),i\in N^+\right\} \tag{8}$$

于是，我们有下面的定理和证明。

定理 1　对于一个给定的信源 S，其均值$\mu_P<\infty$，则对于任意的 $l\in N^+$有

$$\mathrm{LB_G}(l,\mu_P)\le\overline{L_P(G_l)}\le\mathrm{UB_G}(l,\mu_P) \tag{9}$$

式中

$$\mathrm{LB_G}(l,\mu_P)=\lceil\log l\rceil+\frac{\mu_P-2^{\lceil\log l\rceil}}{l}+1 \tag{10}$$

$$\mathrm{UB_G}(l,\mu_P)=\lceil\log l\rceil+\frac{\mu_P-2^{\lceil\log l\rceil}-1}{l}+2 \tag{11}$$

下界 $\mathrm{LB_G}(l,\mu_P)$ 在

$$l_0=2^{\lfloor\log\mu_P\rfloor} \tag{12}$$

处取到最小值，所以对于所有的 $l\in N^+$，

$$\overline{L_P(G_l)}\ge\mathrm{LB_G}(l_0,\mu_P)\ge\log\mu_P+\log\left(\frac{e}{\log e}\right) \tag{13}$$

上界 $\mathrm{UB_G}(l,\mu_P)$ 在

$$l_1=2^{\lfloor\log\max\{\mu_P-1,1\}\rfloor} \tag{14}$$

处取得最小值，且

$$\mathrm{UB_G}(l_1,\mu_P)\le\log\max\{\mu_P-1,1\}+2 \tag{15}$$

证明　由 Golomb 码的定义式(8)以及式(7)知道，一个参数为 l 的 Golomb 码，其码字长度为

$$|g_l(i)|=\left\lfloor\frac{i-1}{l}\right\rfloor+1+\lceil\log l\rceil-u(2^{\lceil\log l\rceil}-l-1-(i-1)\bmod l)\text{。}$$

因此，由式(1)，Golomb 码的平均码长为

$$\begin{aligned}&\overline{L_P(G_l)}\\&=\sum_{i\in N^+}p(i)\left[\left\lfloor\frac{i-1}{l}\right\rfloor+1+\lceil\log l\rceil-u(2^{\lceil\log l\rceil}-l-1-(i-1)\bmod l)\right]\\&=\lceil\log l\rceil+\frac{\mu_P-1}{l}+1\\&\quad-\sum_{i\in N^+}p(i)\left[\left\langle\frac{i-1}{l}\right\rangle+u(2^{\lceil\log l\rceil}-l-1-(i-1)\bmod l)\right]\\&=\lceil\log l\rceil+\frac{\mu_P-1}{l}+1-\sum_{j=0}^{l-1}\sum_{\substack{(i-1)\bmod l=j,\\ i\in N^+}}p(i)\left[\frac{j}{l}+u(2^{\lceil\log l\rceil}-l-1-j)\right]\\&=\lceil\log l\rceil+\frac{\mu_P-1}{l}+1-\sum_{j=0}^{l-1}\tilde{p}(j)w(j,l)\end{aligned} \tag{16}$$

式中

$$\tilde{p}(j)=\sum_{\substack{(i-1)\bmod l=j\\ i\in N^+}}p(i),\quad w(j,l)=\frac{j}{l}+u(2^{\lceil\log l\rceil}-l-1-j)$$

因为当 $0\le j\le l-1$ 时，

$$\frac{2^{\lceil\log l\rceil}}{l}-1\le w(j,l)\le\frac{2^{\lceil\log l\rceil}-1}{l}$$

所以结合式(16)便可推出式(9)，式(10)和式(11)。

通过观察，可以注意到如果 $\lceil\log l\rceil=\lceil\log l'\rceil$，并且 $l<l'$，那么 $\mathrm{LB_G}(l,\mu_P)>\mathrm{LB_G}(l',\mu_P)$，所以函数 $\mathrm{LB_G}(l,\mu_P)$ 的最小值只可能取在 $l=2^k$（$k\ge0$）。然后，我们计算如下的差式：

$$\mathrm{LB_G}(2^{k+1},\mu_P)-\mathrm{LB_G}(2^k,\mu_P)=1-\frac{\mu_P}{2^{k+1}}\tag{17}$$

由式(17)知，当 $2^{k+1}\le\mu_P$ 时，$\mathrm{LB_G}(2^{k+1},\mu_P)\le\mathrm{LB_G}(2^k,\mu_P)$，而当 $2^{k+1}>\mu_P$ 时，$\mathrm{LB_G}(2^{k+1},\mu_P)>\mathrm{LB_G}(2^k,\mu_P)$。因此函数 $\mathrm{LB_G}(2^k,\mu_P)$ 的最小值就是取在满足不等式 $2^{k+1}>\mu_P$ 的最小非负整数 k 上。通过解不等式，就可以推出式(12)。于是，对于所有 $l\in N^+$，

$$\overline{L_P(G_l)}\ge\mathrm{LB_G}(l_0,\mu_P)=\log\mu_P+2^{\langle\log\mu_P\rangle}-\langle\log\mu_P\rangle。$$

由微积分的知识知道，函数 $f(x)=2^x-x$（$0\le x<1$）的最小值取在满足 $f'(x)=0$ 的值上，即 $x_0=\log\log e$，所以，

$$\mathrm{LB_G}(l_0,\mu_P)\ge\log\mu_P+2^{\log\log e}-\log\log e=\log\mu_P+\log\left(\frac{e}{\log e}\right),$$

式(13)得证。

与下界函数的证明类似，同理可以证明上界函数在式(14)处取得最小值，而且当 $l=l_1$ 时，

$$\begin{aligned}&\mathrm{UB_G}(l_1,\mu_P)\\&=\lfloor\log\max\{\mu_P-1,1\}\rfloor+\frac{\mu_P-1}{2^{\lfloor\log\max\{\mu_P-1,1\}\rfloor}}+1\\&\le\lfloor\log\max\{\mu_P-1,1\}\rfloor+\frac{\max\{\mu_P-1,1\}}{2^{\lfloor\log\max\{\mu_P-1,1\}\rfloor}}+1\\&=\log\max\{\mu_P-1,1\}+2^{\langle\log\max\{\mu_P-1,1\}\rangle}-\langle\log\max\{\mu_P-1,1\}\rangle+1\\&\overset{(a)}{\le}\log\max\{\mu_P-1,1\}+2\end{aligned}$$

式中(a)利用了不等式 $2^x\le x+1$（$0\le x\le1$），从而式(15)得证。

证毕

定理 1 不仅给出了具有有限均值的任意分布下 Golomb 码平均码字长度的最紧的上下界，而且还分别给出了 Golomb 码达到最低上下界的参数公式，特别是式(14)，它给出了在已知信源均值的条件下 Golomb 码参数的最优选择准则，这对于实际的应用是非常有指导意义的。从式(13)和式(15)来看，在任意的有限均值分布下，Golomb 码的最优平均码长大约近似为 $\log\mu_P$，所以 Golomb 码的最优性能几乎完全由信源的均值决定。当然，由文献[10]的 Lemma 1 知道，Golomb 码并不是通用码，所以在第 3 节中，我们将以 Golomb 码为基础构造一类带参数的通用码。

3 扩展的 γ 码

Elias γ 码是 Elias 在文献[9]中提出的一类通用码，本节将以 Golomb 码为基础推广 Elias γ 码。我们首先来看一下这种码的具体形式，这里将考察 γ 码的一个变种，如下：$\gamma(i)=(\Lambda(i))_1\,[i]_2$。可以注意到，$\gamma$ 码的一个特点就是其码字中包含着一个一元码，即参数为 1 的 Golomb 码，所以可以很自然地想到将参数为 1 的 Golomb 码替换为参数为 l 的 Golomb 码，于是我们有以下扩展的 γ 码：

$$\Gamma_l=\left\{\gamma_l(i):\gamma_l(i)=g_l(\Lambda(i))[i]_2,i\in N^+\right\}\tag{18}$$

扩展 γ 码的一个子类（即 l 取值为 2 的幂次）在文献[11]中从其他角度被提出过。

首先分析扩展 γ 码在任意分布下的性能，于是有如下定理。

定理 2 对于一个给定的信源 S，定义其取整对数均值为

$$\nu_P\equiv E_P[\lambda(i)]=\sum_{i\in N^+}p(i)\lfloor\log i\rfloor\tag{19}$$

并且 $\nu_P<\infty$，则对于任意的 $l\in N^+$ 有

$$\mathrm{LB_\Gamma}(l,\nu_P)\le\overline{L_P(\Gamma_l)}\le\mathrm{UB_\Gamma}(l,\nu_P)\tag{20}$$

式中

$$\mathrm{LB_\Gamma}(l,\nu_P)=(1+\frac{1}{l})\nu_P+\lceil\log l\rceil-\frac{2^{\lceil\log l\rceil}-1}{l}+1\tag{21}$$

$$\mathrm{UB_\Gamma}(l,\nu_P)=(1+\frac{1}{l})\nu_P+\lceil\log l\rceil-\frac{2^{\lceil\log l\rceil}}{l}+2\tag{22}$$

下界 $\mathrm{LB_\Gamma}(l,\nu_P)$ 在

$$l_0'=2^{\lfloor\log(\nu_P+1)\rfloor}\tag{23}$$

处取到最小值，所以对于所有的 $l\in N^+$，

$$\overline{L_P(\Gamma_l)}\ge\mathrm{LB_\Gamma}(l_0',\nu_P)\ge\nu_P+\log(\nu_P+1)+\log\left(\frac{e}{\log e}\right)\tag{24}$$

上界 $\mathrm{UB_\Gamma}(l,\nu_P)$ 在

$$l_1'=2^{\lfloor\log\max\{\nu_P,1\}\rfloor}\tag{25}$$

处取到最小值，且当 $l=l_1'$ 时，

$$\mathrm{UB_\Gamma}(l_1',\nu_P)\le\nu_P+\log\max\{\nu_P,1\}+2\tag{26}$$

证明 由扩展 γ 码的定义式(18)可以计算得到其码字长度为 $|\gamma_l(i)|=|g_l(\Lambda(i))|+\lambda(i)$，所以，由式(1)，参数为 l 的扩展 γ 码的平均码长为

$$\overline{L_P(\Gamma_l)} = \sum_{i\in N^+} p(i)[|g_l(\Lambda(i))| + \lambda(i)]$$
$$\overset{(a)}{=} \nu_P + \sum_{j\in N^+}\sum_{\substack{\Lambda(i)=j,\\ i\in N^+}} p(i)|g_l(j)| = \nu_P + \overline{L_{P'}(G_l)} \tag{27}$$

式中概率分布 P' 定义为

$$p'(j) = \sum_{\Lambda(i)=j, i\in N^+} p(i)$$

而步骤(a)利用了式(19)以及求和式的恒等变换。又因为

$$\mu_{P'} = \sum_{j\in N^+} jp'(j) = \sum_{i\in N^+} p(i)\Lambda(i) \overset{(a)}{=} \nu_P + 1 \tag{28}$$

式中(a)利用了定义式(5)和式(19)，所以，结合式(27)，式(28)和定理 1 的结论，就可以依次推出式(20)～式(26)。　证毕

进一步，如果信源 $S\in\mathcal{S}_B$，我们有以下的推论。

推论 1　如果 $S\in\mathcal{S}_B$，那么对于任意的 $l\in N^+$有

$$\overline{L_P(\Gamma_l)} \le (1+\frac{1}{l})H(P) + \lceil \log l\rceil - \frac{2^{\lceil \log l\rceil}}{l} + 2 \tag{29}$$

而且

$$\overline{L_P(\Gamma_{l_1'})} \le H(P) + \log\max\{H(P),1\} + 2 \tag{30}$$

式中 l_1' 由式(25)定义。

证明　因为 $S\in\mathcal{S}_B$，所以由文献[12]的 Wyner 不等式知，

$$\nu_P = E_P[\lambda(i)] \le E_P[\log i] \le H(P) \tag{31}$$

由式(22)，式(26)和式(31)便可分别推出式(29)和式(30)。

证毕

根据式(3)和式(29)可以知道扩展 γ 码是一类通用码，当 $H(P)$趋向于无穷时，其平均码长大约为$(1+1/l)H(P)$，因为 $l\ge 1$，所以其系数可以随 l 在 1 到 2 之间变动。具有类似性质的带参数码有文献[13]提出的基于斐波那契（Fibonacci）计数系统的编码方案，以及文献[14,15]提出的基于 Flag 策略的两类编码方案。但是斐波那契码的编解码算法较复杂，而基于 Flag 策略的码字长度并不是单调不增的，所以相比这 3 类方案，虽然扩展 γ 码不具有自同步的鲁棒（Robust）码特性，但对于一般的数据压缩应用却是首选的方案。

4　一个低复杂性的通用数据压缩框架

本节将以 Golomb 码和扩展 γ 码为基础，给出一个低复杂性的基于整数编码的通用数据压缩框架，这是一个两阶段的框架系统（如图 1）。首先，对数据进行预处理，其目的就是将不同的数据类型都变换为一个以正整数为字符集而概率分布为近似递减的信源。因此，对于不同的数据类型会采用不同的预处理技术，如声音、图形等的连续信号数据，可以采用预测技术将信号变换为误差信号，再通过由整数到正整数的一个变换，就可以得到概率分布近似递减的信源（如文献[2]），又如一般的文本数据，可以采用 Burrows-Wheeler 变换（BWT）[16]和 Move-to-Front（MTF）变换[17]对数据进行处理，也可以得到概率分布近似递减的信源。然后，将预处理后的数据编码为 Golomb 码或扩展 γ 码，其参数的选择由μ_P或ν_P来确定。在具体实现中，这两个统计值可以用逐个样本的自适应模式去估计，也可以用逐块样本的非自适应模式去计算，然后将参数写在每块数据的头部。

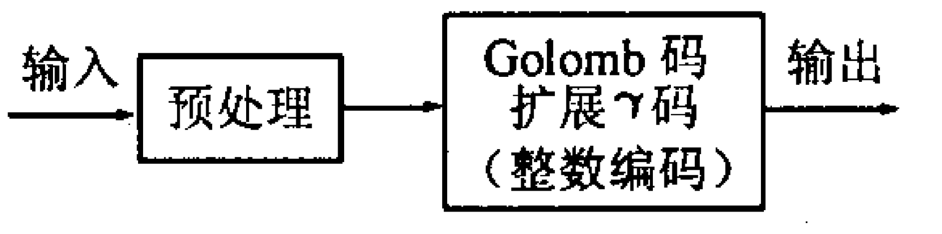

图 1　基于整数编码的通用数据压缩框架

我们以此框架为基础，通过在预处理阶段采用 MTF 变换，构建了一个简单的通用数据压缩样例系统，以此说明 Golomb 码和 γ 码的实际应用价值。MTF 变换的主要原理就是通过自适应的方式对各字符进行重新排列，使高概率字符趋向于排在低概率字符的前面，然后输出重新排列后的列表索引。文献[17, 18]中的理论分析表明 MTF 配合渐近最优的整数编码，同时结合字符扩展技术，可以达到渐近于信源熵率的性能。在该系统中，我们让 Golomb 码和 γ 码分别工作于自适应与非自适应方式，其对于 Calgary Corpus 数据测试集的压缩结果如表 1，表中也列出了数据测试集的平均一阶熵。

表 1　Calgary Corpus 的压缩结果

压缩方案	平均码字长度（bit/symbol）
数据测试集平均一阶熵	4.909
Golomb 码（非自适应方式）	5.484
Golomb 码（自适应方式）	5.319
扩展 γ 码（非自适应方式）	6.032
扩展 γ 码（自适应方式）	5.816

5　结论

本文通过对任意分布下 Golomb 码和扩展 γ 码的性能分析，实际上给出了两类包含最优参数选择准则的自适应编码方案。不同于霍夫曼编码要求已知信源的概率分布，这两类码在参数μ_P和ν_P确定时，其最优的参数和最优的平均码长就已经基本确定。特别是扩展 γ 码，当已知ν_P时，它可以达到渐近最优的性能。因此，这两类码都十分适合应用于要求低复杂性的自适应数据压缩方案，结合最后我们提出的一个低复杂性的基于整数编码的通用数据压缩框架，实际上给出了

一整套行之有效的压缩系统设计方法，这对于实际压缩系统的设计具有重要的指导意义。

参 考 文 献

[1] Golomb S W. Run-length encodings [J]. *IEEE Trans. on Info. Theory*, 1966, 12(3): 399 – 401.

[2] Weinberger M J, Seroussi G, Sapiro G. The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS [J]. *IEEE Trans. on Image Processing*, 2000, 9(8): 1309 – 1324.

[3] Rice R F. Some practical universal noiseless coding techniques—Part III [R]. Technical Report JPL-91-3, Jet Propulsion Laboratory, Pasadena, CA, 1991.

[4] Gallager R G, Voorhis D C V. Optimal source codes for geometrically distributed integer alphabets [J]. *IEEE Trans. on Info. Theory*, 1975, 21(2): 228 – 230.

[5] Szpankowski W. Asymptotic average redundancy of Huffman (and other) block codes [J]. *IEEE Trans. on Info. Theory*, 2000, 46(7): 2434 – 2443.

[6] Merhav N, Seroussi G, Weinberger M J. Optimal prefix codes for sources with two-sided geometric distributions [J]. *IEEE Trans. on Info. Theory*, 2000, 46(1): 121 – 135.

[7] Howard P G. The design and analysis of efficient lossless data compression systems [D]. Rhode Island: Brown University, 1993.

[8] Fenwick P. Punctured Elias codes for variable-length coding of the integers [R]. Technical Report 137, Dept of Computer Science, The University of Auckland, New Zealand, December 1996.

[9] Elias P. Universal codeword sets and representations of the integers [J]. *IEEE Trans. on Info. Theory*, 1975, 21(2): 194 – 203.

[10] Lakshmanan K B. On universal codeword sets [J]. *IEEE Trans. on Info. Theory*, 1981, 27(5): 659 – 662.

[11] Amemiya T, Yamamoto H. A new class of the universal representation for the positive integers [J]. *IEICE Trans. on Fundamentals*, 1993, E76-A(3): 447 – 452.

[12] Wyner A D. An upper bound on the entropy series [J]. *Information and Control*, 1972, 20(2): 176 – 181.

[13] Apostolico A, Fraenkel A S. Robust transmission of unbounded strings using Fibonacci representations [J]. *IEEE Trans. on Info. Theory*, 1987, 33(2): 238 – 245.

[14] Wang M. Almost asymptotically optimal flag encoding of the integers[J]. *IEEE Trans. on Info. Theory*, 1988, 34(2): 324 – 326.

[15] Yamamoto H, Ochi H. A new asymptotically optimal code for the positive integers [J]. *IEEE Trans. on Info. Theory*, 1991, 37(5): 1420 – 1429.

[16] Burrows M, Wheeler D J. A block-sorting lossless data compression algorithm [R]. Technical Report SRC 124, Digital System Research Center, Palo Alto, CA, 1994.

[17] Bentley J L, Sleator D D, Tarjan R E, *et al*.. A locally adaptive data compression scheme[J]. *Commun. ACM*, 1986, 29(4): 320 – 330.

[18] Muramatsu J. On the performance of recency-rank and block-sorting universal lossless data compression algorithms[J]. *IEEE Trans. on Info. Theory*, 2002, 48(9): 2621 – 2625.

杨胜天：　男，1976 年生，博士生，研究方向为信息论与编码理论.

仇佩亮：　男，1944 年生，教授，博士生导师，研究方向为信息论与编码、通信理论与技术.