关于二元等重码的最大码字数

夏树涛
(清华大学深圳研究生院,广东深圳 518055)

摘 要: 本文利用 Johnson Schemes理论研究了二元等重码及其最大码字数问题. 在 Delsarte的 associate schemes理论中, Q-变换被引入以研究二元等重码的距离分布. 首先,本文研究了等重码距离分布的 Q-变换; 然后,通过使用 Q-变换的性质,我们研究了二元等重码的最大码字数问题并得到码字数的一个新的上界, 该上界在形式上类似于纠错码理论中的 Grey-Rankin界,并且在某些情况下优于已知的结果.

关键词: 二元等重码; 最大码字数; 距离分布; Johnson Schemes; Q-变换
中图分类号: TN911.2 文献标识码: A 文章编号: 0372-2112 (2006) 09-1613-03

On the Maximum Number of Codewords of Binary Constant Weight Codes

XIA Shu-tao
(*The Graduate School at Shenzhen, Tsinghua University, Shenzhen, Guangdong* 518055, *China*)

Abstract The problems of maximum number of codewords for binary constant weight codes are studied by employing the theory of Johnson Schemes. In Delsarte's association schemes theory, Q-transform were introduced to study the distance distributions of binary constant weight codes. First, we study the Q-transforms of distance distributions of binary constant weight codes. Then, by using the properties of Q-transforms, we obtain a new upper bound of number of codewords for binary constant weight codes. This bound is similar to Grey-Rankin bound in error-correcting codes theory in form and improves previously known results in certain cases.

Key words binary constant weight codes; maximum number of codewords; distance distribution; Johnson schemes; Q-transform

1 引言

二元等重码在纠错码理论中占有重要的地位, 在理论和应用都具有重要意义. 二元等重码的研究成果非常丰富,在计算机和通信系统中有着很多的应用, 如 ARQ 差错控制系统(请参阅文[1]、[2]及其参考文献). 记 $V_n = \{0, 1\}^n$ 为二元 n维向量空间, 称 C是一个二元 (n, M)码, 若 C是 V_n的一个包含 M 个向量的子集. 若 C的所有码字具有相同的重量 w, 则称 C为二元 (n, M, w)等重码. 记 $A(n, d)$和 $A(n, d, w)$分别为极小距离大于等于 d的二元码和二元等重码的最大码字数目. 确定 $A(n, d)$和 $A(n, d, w)$的精确值或好的上下界是编码理论的主要问题之一, 至今尚未完全解决. 本文利用 Johnson Schemes理论研究了二元等重码及其最大码字数问题. 首先, 我们研究了等重码距离分布的 Q-变换; 然后, 通过使用 Q-变换分布的性质, 我们研究了二元等重码的最大码字数问题并得到码字数的一个新的上界, 该上界在形式上类似于纠错码理论中的 Grey-Rankin界[3], 并且在某些情况下优于文[4]中的结果.

2 二元等重码的距离分布

设 C为二元 (n, M, d, w)等重码, 其中 n为码长, M为码字个数, d为极小距离, w为码字的重量. 显然, $d \leqslant 2w$, 而且任何两个码字之间的距离一定是偶数. C的距离分布定义为

$$A_i = \frac{1}{M} \left| \{(a, b): a, b \in C, d_H(a, b) = 2i\} \right|, \quad i = 0, 1, \cdots, w \tag{1}$$

容易知道

$$A_0 = 1, \sum_{i=0}^{w} A_i = M, \tag{2}$$

$$\sum_{i=0}^{w} 2iA_i = \frac{1}{M} \sum_{a, b \in C} d_H(a, b) \tag{3}$$

收稿日期: 2005-03-25; 修回日期: 2006-06-13
基金项目: 国家自然科学基金 (No. 60402031)

$$\sum_{i=0}^{w} 4^i A_i = \frac{1}{M}\sum_{b\in C} d_H^2(a, b) \tag{4}$$

距离分布的 Q变换分布 B_j 定义为

$$B_j = \frac{1}{M}\sum_{i=0}^{w} Q_j(i)A_i,\ j = 0, 1, \cdots, w \tag{5}$$

其中

$$Q_j(x) = \left[\binom{n}{j} - \binom{n}{j-1}\right]\sum_{k=0}^{j}(-1)^k \frac{\binom{j}{k}\binom{n+1-j}{k}}{\binom{w}{k}\binom{n-w}{k}}\binom{x}{k} \tag{6}$$

称为对偶 Hahn多项式. 令

$$u_j = \binom{n}{j} - \binom{n}{j-1} \tag{7}$$

$$v_i = \binom{w}{i}\binom{n-w}{i} \tag{8}$$

对偶 Hahn多项式具有以下的性质 (请参见文[3]):

$$Q_0(x) = 1,\ Q_j(0) = u_j,\ Q_j(i) = \frac{u_j}{v_i}E_i(j) \tag{9}$$

$$E_i(x) = \sum_{k=0}^{i}(-1)^k \binom{x}{k}\binom{w-x}{i-k}\binom{n-w-x}{i-k} \tag{10}$$

Delsarte[5]建立的 Association Schemes理论在编码理论中具有重要应用, 其中 Johnson Schemes针对二元等重码 (详见文[3]), 注意到 $E_k(i)$和 $Q_k(i)$是 Johnson Schemes的两个特征值.

引理 1[3,5] $B_0 = 1, B_k \geqslant 0,\ k = 1, 2, \cdots, w.$

3 主要结果

设 C 为二元 (n, M, d, w)等重码, 其中 n为码长, M 为码字个数, d 为极小距离, w 为码字的重量. 由式 (10)不难得到,

$$E_i(1) = v_i\left[1 - \frac{in}{w(n-w)}\right] \tag{11}$$

由式 (7)知 $u_1 = n-1$. 所以, 结合式 (2), (3), (5), (9), 有

$$\begin{aligned} B_1 &= \frac{1}{M}\sum_{i=0}^{w}\frac{u_1}{v_i}E_i(1)A_i \\ &= \frac{n-1}{M}\sum_{i=0}^{w}\left[1 - \frac{in}{w(n-w)}\right]A_i \\ &= (n-1) - \frac{n(n-1)}{2v(n-w)}\left(\frac{1}{M^2}\sum_{a}\sum_{b\in C} d_H(a, b)\right) \end{aligned} \tag{12}$$

即

$$\frac{1}{M^2}\sum_{a}\sum_{b\in C} d_H(a, b) = (n-1-B_1)\frac{2v(n-w)}{n(n-1)}.$$

因此, 由 $B_1 \geqslant 0$和 $d_H(a, b) \geqslant d$ 知 $\frac{M-1}{M}d \leqslant \frac{2v(n-w)}{n}$, 稍作变形即得下面的 Johnson界.

命题 1[3] 对于正整数 n, d, w, 若 $nd - 2vn + 2v^2 > 0$, 则 $A(n, d, w) \leqslant \frac{nd}{nd - 2vn + 2v^2}$.

这样, 我们证明了 $B_1 \geqslant 0$可以直接得到著名的 Johnson界. 下面, 我们通过 $B_1 \geqslant 0$和 $B_2 \geqslant 0$推导文[4]中的主要结果. 由式 (3)~ (9), 不难得到

$$\begin{aligned} B_2 &= \frac{1}{M}\sum_{i=0}^{w}\frac{u_2}{v_i}E_i(2)A_i = \frac{n(n-3)}{2M} \\ &\quad \cdot \sum_{i=0}^{w}\left[1 + \frac{i^2(n-1)(n-2) + i(n-1)(n-2vn+2v^2)}{w(w-1)(n-w)(n-w-1)}\right]A_i \\ &= \frac{n(n-3)}{2} + \frac{n(n-1)(n-2)(n-3)}{8v(w-1)(n-w)(n-w-1)}\left(\frac{1}{M^2}\sum_{a}\sum_{b\in C} d_H^2(a, b)\right) \\ &\quad - \frac{n(n-1)(n-3)[n(n-2) - (n-2v)^2]}{8v(w-1)(n-w)(n-w-1)}\left(\frac{1}{M^2}\sum_{a}\sum_{b\in C} d_H(a, b)\right) \end{aligned} \tag{13}$$

利用式 (12)和 (13), 我们可以用 B_1 和 B_2 解出 $\sum_{a}\sum_{b\in C} d_H(a, b)/M^2$和 $\sum_{a}\sum_{b\in C} d_H^2(a, b)/M^2$, 由此不难算出

$$\begin{aligned} &\frac{1}{M^2}\sum_{a}\sum_{b\in C} d_H(a, b)\left[n - d_H(a, b)\right] \\ &= 2v(n-w) - \frac{4v^2(n-w)^2}{n(n-1)} - B_1\frac{2v(n-w)(n-2v)^2}{n(n-1)(n-2)} \\ &\quad - B_2\frac{8v(w-1)(n-w)(n-w-1)}{n(n-1)(n-2)(n-3)} \end{aligned} \tag{14}$$

假设 $n - d \geqslant d_H(a, b) \geqslant d$, 则

$$\frac{1}{M^2}\sum_{a}\sum_{b\in C} d_H(a, b)\left[n - d_H(a, b)\right] \geqslant \frac{M-1}{M}d(n-d) \tag{15}$$

这样, 由 $B_1 \geqslant 0, B_2 \geqslant 0$, 式 (14)和 (15)可得到下面的不等式 (16)

$$\frac{M-1}{M}d(n-d) \leqslant 2v(n-w) - \frac{4v^2(n-w)^2}{n(n-1)} \tag{16}$$

由不等式 (16)可立刻得到文[4]定理 1.1的上界, 即命题 2

命题 2[4] 若二元 (n, M, w)等重码 C满足 $\forall a \neq b \in C$, $d \leqslant d_H(a, b) \leqslant n-d$, 则当 $S_0 = d(n-d)n(n-1) + 2v(n-w)n(n-1) + 4v^2(n-w)^2 > 0$时 $M \leqslant d(n-d)n(n-1)/S_0$ (17)

上面我们通过 $B_1 \geqslant 0$和 $B_2 \geqslant 0$给出了命题 2的另外一个证明. 现在, 我们开始推导新的上界, 该上界不需要使用 $B_1 \geqslant 0$. 将式 (13)变形可得

$$\begin{aligned} &\frac{1}{M^2}\sum_{a}\sum_{b\in C} d_H(a, b)\left[n - \frac{(n-2v)^2}{n-2} - d_H(a, b)\right] \\ &= \frac{4w(w-1)(n-w)(n-w-1)}{(n-1)(n-2)} \\ &\quad - B_2\frac{8w(w-1)(n-w)(n-w-1)}{n(n-1)(n-2)(n-3)} \end{aligned} \tag{18}$$

当 $n - \frac{(n-2v)^2}{n-2} \geqslant d_H(a, b) \geqslant d$时,

$$\begin{aligned} &\frac{1}{M^2}\sum_{a}\sum_{b\in C} d_H(a, b)\left[n - \frac{(n-2v)^2}{n-2} - d_H(a, b)\right] \\ &\geqslant \frac{M-1}{M}d\left[n - \frac{(n-2v)^2}{n-2} - d\right] \end{aligned} \tag{19}$$

这样, 由 $B_2 \geqslant 0$, 式 (18)和 (19)可得到下面的不等式 (20)

$$\frac{M-1}{M}d\left[n-\frac{(n-2w)^2}{n-2}-d\right]$$
$$\leqslant\frac{4w(w-1)(n-w)(n-w-1)}{(n-1)(n-2)} \quad (20)$$

由不等式(20)可立刻得到以下定理.

定理 1　若二元(n, M, d, w)等重码 C 满足 $\forall a\neq b\in C$, $d\leqslant d_H(a, b)\leqslant n-\frac{(n-2w)^2}{n-2}-d$, 则当 $S>T$ 时,

$$M\leqslant\frac{S}{S-T} \quad (21)$$

其中

$$S=d\left[n-\frac{(n-2w)^2}{n-2}-d\right] \quad (22)$$

$$T=\frac{4w(w-1)(n-w)(n-w-1)}{(n-1)(n-2)} \quad (23)$$

命题 2和定理 1在形式上都有些类似于二元码的 Grey-Rankin界[3], 只不过处理的对象不是普通二元码而是二元等重码. 下面, 我们比较一下定理 1和命题 2中的两个上界. 由式(14)~(16)可知, $B_1=B_2=0$是达到命题 2的上界的必要条件; 而由式(18)~(20)可知, $B_2=0$是达到定理 1的上界的必要条件, 并不要求 $B_1=0$, 所以我们有理由相信定理 1在某些情况下要优于命题 2. 事实上, 不难验证以下结论: (1)当 $n=2w$ 时, 定理 1与命题 2完全等价; (2)假定两个上界的条件都满足, 当 $d<\frac{2w(n-w)}{n-1}$时, 定理 1的上界优于命题 2的上界; (3)假定两个上界的条件都满足, 当 $d=\frac{2w(n-w)}{n-1}$时, 定理 1的上界等价于命题 2的上界; (4)假定两个上界的条件都满足, 当 $d>\frac{2w(n-w)}{n-1}$时, 定理 1的上界劣于命题 2的上界.

例: 令 $n=13$, $d=w=6$, 由于二元等重码的码字间距离为偶数, 故容易知道命题 2与定理 1所要求的假设条件是相同的, 命题 2表明 $M\leqslant 13$; 而定理 1表明 $M\leqslant 12.67$, 即 $M\leqslant 12$. 另外, 令 C_0 由以下码字组成{1111110000000, 1110001110000, 1110000001110, 1101001001001, 1100100100101, 1100010010011, 1011000100011, 1010100011001, 1010011000101, 0111000010101, 0110101000011, 0110010101001}. 容易验证 C_0 是达到定理 1上界的二元等重码. 事实上, 稍加分析可知 C_0 还是等距码, 而且是最优等距等重码.

关于达到 Grey-Rankin界的二元码已取得了一些重要进展(请参阅文[6]), 而达到定理 1上界的二元等重码是否有类似的结果? 这是一个值得探讨的问题.

参考文献:

[1] Wang X M, Yang Y X. On the undetected error probability of nonlinear binary constant weight codes[J]. IEEE Trans Commu, 1994, 42(7): 2390-2393.

[2] Fu F W, Xia S T. Binary constant weight codes for error detection[J]. IEEE Trans Inform Theory, 1998, 44(3): 1294-1299.

[3] Macwilliams F J, Sloane N J A. The Theory of Error-correcting Codes[M]. North-Holland: Elsevier Science Publishers, 1977. Ch21, 525, 544.

[4] Fu F W, Shen S Y. An upper bound for binary constant weight codes [J]. Journal of Statistical Planning and Inference, 2001, 94: 197-203.

[5] Delsarte P. Bounds for unrestricted codes, by linear programming[J]. Philips Research Reports, 1972, 27: 272-289.

[6] Mcguire G. Quasi-symmetric designs and codes meeting the Grey-Rankin bound[J]. Journal of Combinatory Theory Series A, 1997, 78: 280-291.

作者简介:

夏树涛　男, 1972年 12月生于黑龙江省, 1997年毕业于南开大学数学学院, 获理学博士学位. 1997年 9月-1998年 9月期间在香港中文大学讯息工程系作访问学者, 现为清华大学深圳研究生院信息学部副教授, 主要从事信道编码和网络安全等方向的教学与科研工作, 目前负责或参加国家自然科学基金、973等多项课题, 在国际国内学术期刊及国际会议论文集上发表学术论文二十余篇.

E-mail: xiasf@sz.tsinghua.edu.cn