

# 信息论作业 4

史泽宇

2020 年 3 月 19 日

## 题目 1

表 1: 二元码

字母	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$
概率 $P$	0.16	0.14	0.13	0.12	0.1	0.09	0.08	0.07	0.06	0.05
编码长度 $l$	3	3	3	3	3	3	4	4	4	4
编码	111	101	100	011	001	000	1101	1100	0101	0100

1. 平均编码长度为

$$average(l) = \sum_{i=1}^{10} P(a_i)l(a_i) \quad (1)$$

$$= 3.26 \quad (2)$$

编码效率为

$$\eta = \frac{H(A)}{average(l)} \quad (3)$$

$$= \frac{-\sum_{i=1}^{10} P(a_i) \log_2 P(a_i)}{average(l)} \quad (4)$$

$$\approx \frac{3.2344}{3.26} \quad (5)$$

$$\approx 0.9921 \quad (6)$$

表 2: 三元码

字母	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$
概率 $P$	0.16	0.14	0.13	0.12	0.1	0.09	0.08	0.07	0.06	0.05
编码长度 $l$	2	2	2	2	2	2	2	2	3	3
编码	22	21	20	12	10	02	01	00	111	110

2. 第一次合并节点数为

$$((10 - 2) \bmod (3 - 1)) + 2 \quad (7)$$

$$= 2 \quad (8)$$

平均编码长度为

$$average(l) = \sum_{i=1}^{10} P(a_i)l(a_i) \quad (9)$$

$$= 2.11 \quad (10)$$

编码效率为

$$\eta = \frac{H_3(A)}{average(l)} \quad (11)$$

$$= \frac{-\sum_{i=1}^{10} P(a_i) \log_3 P(a_i)}{average(l)} \quad (12)$$

$$\approx \frac{2.0407}{2.11} \quad (13)$$

$$\approx 0.9671 \quad (14)$$

## 题目 2

表 3: 二源码

字母	$a_1$	$a_2$	$a_3$
概率 $P$	0.5	0.3	0.2
编码长度 $l$	1	2	2
编码	0	11	10

1. 平均编码长度为

$$average(l) = \sum_{i=1}^3 P(a_i)l(a_i) \quad (15)$$

$$= 1.5 \quad (16)$$

编码效率为

$$\eta = \frac{H(A)}{average(l)} \quad (17)$$

$$= \frac{-\sum_{i=1}^3 P(a_i) \log_2 P(a_i)}{average(l)} \quad (18)$$

$$\approx \frac{1.4855}{1.5} \quad (19)$$

$$\approx 0.9903 \quad (20)$$

表 4: 二源码

字母	$a_1a_1$	$a_1a_2$	$a_1a_3$	$a_2a_1$	$a_2a_2$	$a_2a_3$	$a_3a_1$	$a_3a_2$	$a_3a_3$
概率 $P$	0.25	0.15	0.1	0.15	0.09	0.06	0.1	0.06	0.04
编码长度 $l$	2	3	3	3	4	4	3	4	4
编码	10	110	001	111	0111	0101	000	0110	0100

2. 平均编码长度为

$$average(l) = \sum_{i=1}^3 \sum_{j=1}^3 P(a_i a_j) l(a_i a_j) \quad (21)$$

$$= 3.0 \quad (22)$$

编码效率为

$$\eta = \frac{H(A)}{average(l)} \quad (23)$$

$$= \frac{-\sum_{i=1}^3 \sum_{j=1}^3 P(a_i a_j) \log_2 P(a_i a_j)}{average(l)} \quad (24)$$

$$\approx \frac{2.9710}{3.0} \quad (25)$$

$$\approx 0.9903 \quad (26)$$

表 5: 二源码

字母	$a_1a_1a_1$	$a_1a_1a_2$	$a_1a_1a_3$	$a_1a_2a_1$	$a_1a_2a_2$	$a_1a_2a_3$	$a_1a_3a_1$	$a_1a_3a_2$	$a_1a_3a_3$
概率 $P$	0.125	0.075	0.05	0.075	0.045	0.03	0.05	0.03	0.02
编码长度 $l$	3	4	4	4	4	5	4	5	6
编码	100	1100	0011	1101	0000	01101	0010	01111	111100
字母	$a_2a_1a_1$	$a_2a_1a_2$	$a_2a_1a_3$	$a_2a_2a_1$	$a_2a_2a_2$	$a_2a_2a_3$	$a_2a_3a_1$	$a_2a_3a_2$	$a_2a_3a_3$
概率 $P$	0.075	0.045	0.03	0.045	0.027	0.018	0.03	0.018	0.012
编码长度 $l$	4	4	5	5	5	6	5	6	7
编码	1011	0001	01110	11111	01010	101010	10100	111000	1111011
字母	$a_3a_1a_1$	$a_3a_1a_2$	$a_3a_1a_3$	$a_3a_2a_1$	$a_3a_2a_2$	$a_3a_2a_3$	$a_3a_3a_1$	$a_3a_3a_2$	$a_3a_3a_3$
概率 $P$	0.05	0.03	0.02	0.03	0.018	0.012	0.02	0.012	0.008
编码长度 $l$	4	5	6	5	6	7	6	7	7
编码	0100	01100	111011	01011	101011	1111010	111001	1110101	1110100

3. 平均编码长度为

$$average(l) = \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 P(a_i a_j a_k) l(a_i a_j a_k) \quad (27)$$

$$= 4.4870 \quad (28)$$

编码效率为

$$\eta = \frac{H(A)}{\text{average}(l)} \quad (29)$$

$$= \frac{-\sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 P(a_i a_j a_k) \log_2 P(a_i a_j a_k)}{\text{average}(l)} \quad (30)$$

$$\approx \frac{4.4564}{\text{average}(l)} \quad (31)$$

$$\approx 0.9932 \quad (32)$$

随着字母数量的增多，平均编码长度与信息熵应该同步增加，并使得编码效率不变。在第一小问与第二小问中，这显然成立。可是在第三小问中，发现编码效率改变了，这是为什么呢。