2019/2/25 課堂作業1

1. 針對Dataset資料夾內的檔案利用7-zip(V19.00 x64)軟體將其壓成.zip及7z格式，將結果製成表格並計算壓縮比(取到小數第2位)，並分析所得結果。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 檔案 | 原始大小(Bytes) | Zip  (Bytes) | 壓縮比 | 7z  (Bytes) | 壓縮比 |
| world95.txt | 3,005,020 | 841,765 | 3.57 | 593,772 | 5.06 |
| water\_lilies.bmp | 1,440,054 | 1,021,711 | 1.41 | 881,343 | 1.63 |
| EXCEL.EXE | 34,482,256 | 15,315,972 | 2.25 | 11,689,865 | 2.95 |
| Bach-PartitaEmajor-44kHz-Stereo-16bit.wav | 9,580,594 | 8,859,205 | 1.08 | 6,881,670 | 1.39 |
| bridge-close\_cif.yuv  (4:2:0 YUV , 2000 Frames) | 304,128,000 | 177,771,701 | 1.71 | 154,500,920 | 1.97 |
| 720p50\_parkrun\_ter.y4m | 696,732,659 | 506,145,734 | 1.38 | 456,183,711 | 1.53 |

* 從上表可發現壓成7z格式會比.zip格式得到較好的壓縮比。
* 同一種壓縮方法其壓縮比會因不同類型的原始檔而有相當大的變化。其中以文字類型的檔案world95.txt有最好的結果。

1. 為一筆資料做資料壓縮的同時，也等於為該筆資料做資訊保密，對不對，請加以說明。

Ans: 對。因為資料壓縮的過程會對原始資料重新編碼，故其編碼方式已與原先編碼方式不同，因此亦會有資訊保密的效果。

1. 有一張3位元灰階影像，其各灰階值的出現機率及採用變異長度編碼的長度如下表所示，求其平均編碼長度及壓縮比。

|  |  |  |
| --- | --- | --- |
| 灰階值 | 機率 | 編碼長度 |
| 0 | 0.19 | 2 |
| 1 | 0.25 | 2 |
| 2 | 0.21 | 2 |
| 3 | 0.16 | 3 |
| 4 | 0.08 | 4 |
| 5 | 0.06 | 5 |
| 6 | 0.03 | 6 |
| 7 | 0.02 | 6 |

Ans: 平均編碼長度=2.7，壓縮比=3/2.7=1.111。

1. 從52張撲克牌中抽取一張出來可得到多少位元的資訊？在甚麼情形下會有最大熵值(entropy)？在甚麼情形下會有最小熵值(entropy)？

Ans: 抽取一張出來可得到I(pi)=log2(52)=5.7位元的資訊

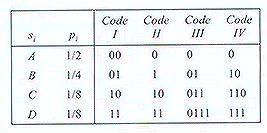
最大熵值發生在每張牌抽中的機率都一樣時，即pi=1/52。

最小熵值發生在其中一張牌抽中的機率是1，其他張抽中的機率是0時。

1. 給定*S*={*s1*, *s2*, *s3*, *s4*, *s5*, *s6,* s7}, p(s1)= p(s2)=1/4，p(s3)= p(s4)= p(s5)=1/8， p(s6)= p(s7)=1/16。求其熵值(entropy)？若有一編碼法可得平均編碼長度2bits。可能嗎？為什麼？

Ans: 熵值H2(S)= 2\*1/4\*log24+3\*1/8\*log28+2\*1/16\*log216 =21/8=2.625 bits。因為平均編碼長度Lavg=2 bits< H2(S)=2.625，故該編碼法不可能存在。

1. 考慮以下四組編碼·（1）哪些碼可唯一解碼，(2)哪些碼不是即時碼？



Ans: （1）可唯一解碼：I、III、IV

（2）不是即時碼：II、III。

評分標準：

* 至少需答對任三題
* 答對三題得基本分70分，每增加一題加10分。