

Zadanie 2 z listy 3 - “Kompresja Danych”

Łukasz Klasieński

4 kwietnia 2020

Zadanie 2

Ustal czy zwiększenie długości słowa kodowego kodu Tunstalla (przy ustalonym alfabecie i prawdopodobieństwach) może spowodować wzrost średniej długości kodu (bps)?

Rozwiązanie

Tak może - podam przykład:

Weźmy następujące dane:

| Σ | p |
|----------|------|
| a | 0.25 |
| b | 0.25 |
| c | 0.25 |
| d | 0.25 |

Wtedy dla $n = 4$ otrzymamy następujące kodowanie:

| Σ | p | kod |
|----------|--------|------|
| aa | 0.0625 | 0000 |
| ab | 0.0625 | 0001 |
| ac | 0.0625 | 0010 |
| ad | 0.0625 | 0011 |
| ba | 0.0625 | 0100 |
| bb | 0.0625 | 0101 |
| bc | 0.0625 | 0110 |
| bd | 0.0625 | 0111 |
| ca | 0.0625 | 1000 |
| cb | 0.0625 | 1001 |
| cc | 0.0625 | 1010 |
| cd | 0.0625 | 1011 |
| da | 0.0625 | 1100 |
| db | 0.0625 | 1101 |
| dc | 0.0625 | 1110 |
| dd | 0.0625 | 1111 |

Oraz średnią długość $= 4/2.0 = 2.0$

Zwiększmy n o 1. Dostajemy następujące prawdopodobieństwa:

| Σ | p |
|----------|--------|
| bb | 0.0625 |
| bc | 0.0625 |
| bd | 0.0625 |

| Σ | p |
|------------|----------|
| <i>ca</i> | 0.0625 |
| <i>cb</i> | 0.0625 |
| <i>cc</i> | 0.0625 |
| <i>cd</i> | 0.0625 |
| <i>da</i> | 0.0625 |
| <i>db</i> | 0.0625 |
| <i>dc</i> | 0.0625 |
| <i>dd</i> | 0.0625 |
| <i>aaa</i> | 0.015625 |
| <i>aab</i> | 0.015625 |
| <i>aac</i> | 0.015625 |
| <i>aad</i> | 0.015625 |
| <i>aba</i> | 0.015625 |
| <i>abb</i> | 0.015625 |
| <i>abc</i> | 0.015625 |
| <i>abd</i> | 0.015625 |
| <i>aca</i> | 0.015625 |
| <i>acb</i> | 0.015625 |
| <i>acc</i> | 0.015625 |
| <i>acd</i> | 0.015625 |
| <i>ada</i> | 0.015625 |
| <i>adb</i> | 0.015625 |
| <i>adc</i> | 0.015625 |
| <i>add</i> | 0.015625 |
| <i>baa</i> | 0.015625 |
| <i>bab</i> | 0.015625 |
| <i>bac</i> | 0.015625 |
| <i>bad</i> | 0.015625 |

Oraz średnią długość $= 5/2.3125 = 2.(162)$. Widać zatem, że dla danych ze zbliżonymi p-p średnia długość niekoniecznie się zmniejsza.