

Információelmélet

Programozási feladatok (laborfeladatok)

Bodó Zalán

Babeş–Bolyai Tudományegyetem
Matematika és Informatika Kar

1. Természetes nyelvek entrópiája
2. Entrópia, mutuális információ, relatív entrópia
3. Huffman- és Shannon–Fano kódok I.
4. Huffman- és Shannon–Fano kódok II.
5. Szótár alapú kódolások
6. Hibajavító kódok
7. ISBN

1. Természetes nyelvek entrópiája
2. Entrópia, mutuális információ, relatív entrópia
3. Huffman- és Shannon–Fano kódok I.
4. Huffman- és Shannon–Fano kódok II.
5. Szótár alapú kódolások
6. Hibajavító kódok
7. ISBN

1. Természetes nyelvek entrópiája
2. Entrópia, mutuális információ, relatív entrópia
3. Huffman- és Shannon–Fano kódok I.
4. Huffman- és Shannon–Fano kódok II.
5. Szótár alapú kódolások
6. Hibajavító kódok
7. ISBN

1. Természetes nyelvek entrópiája

Feladat

Számoljuk ki legalább 2 különböző természetes nyelv (pl. magyar és román) entrópiáját.

- (a) Unigrammokat (azaz szavakat) tekintve.
- (b) Bigrammok esetén.
- (c) Trigrammok esetén.

Szövegekhez használható pl. a *Project Gutenberg* adatbázisa (<https://www.gutenberg.org/>).

1. Természetes nyelvek entrópiája

2. Entrópia, mutuális információ, relatív entrópia

3. Huffman- és Shannon-Fano kódok I.

4. Huffman- és Shannon-Fano kódok II.

5. Szótár alapú kódolások

6. Hibajavító kódok

7. ISBN

2. Entrópia, mutuális információ, relatív entrópia

Feladat

Adott X , Y bináris (Bernoulli) valószínűségi változók, melyek között az $X \rightarrow Y$ kapcsolat áll fenn, és ismerjük a $P(X)$ -et, illetve a $P(Y|X)$ -eket. Írjatok programot (tetszőleges programozási nyelvben), amely kiszámítja az alábbiakat:

- ▶ $H(X), H(Y)$
- ▶ $H(X, Y), H(X|Y), H(Y|X)$
- ▶ $I(X; Y)$
- ▶ $D(P(X)||P(Y))$
- ▶ $D(P(Y)||P(X))$

A program a bemeneteket egy háromsoros szöveges fájlból kapja, ahol a sorokban szereplő értékek a következőket jelentik:

$$P(X = 1)$$

$$P(Y = 1|X = 0)$$

$$P(Y = 1|X = 1)$$

1. Természetes nyelvek entrópiája

2. Entrópia, mutuális információ, relatív entrópia

3. Huffman- és Shannon-Fano kódok I.

4. Huffman- és Shannon-Fano kódok II.

5. Szótár alapú kódolások

6. Hibajavító kódok

7. ISBN

Példa:

0.2

0.1

0.7

=>

$$H(X) = 0.721928$$

$$H(Y) = 0.760168$$

$$H(X,Y) = 1.273383$$

$$H(X|Y) = 0.513215$$

$$H(Y|X) = 0.551455$$

$$I(X;Y) = 0.208713$$

$$D(P(X) || P(Y)) = 0.001720$$

$$D(P(Y) || P(X)) = 0.001761$$

1. Természetes nyelvek
entrópiája

2. Entrópia, mutuális
információ, relatív
entrópia

3. Huffman- és
Shannon-Fano kódok
I.

4. Huffman- és
Shannon-Fano kódok
II.

5. Szótár alapú
kódolások

6. Hibajavító kódok

7. ISBN

3. Huffman- és Shannon–Fano kódok I.

Feladat

Adottak a kódolandó szimbólumok megjelenési valószínűségei (magukat a szimbólumokat nem szükséges ismernünk jelen esetben). Implementáljuk a (bináris és statikus) Huffman- és Shannon–Fano kódolást, határozzuk meg a kódszavakat, illetve azok hosszait, majd számoljuk ki a következőket (tetszőleges programozási nyelvben):

- ▶ A val. változó entrópiája
- ▶ Huffman-kód átlagos kódszóhossza
- ▶ Shannon–Fano-kód átlagos kódszóhossza

A program a bemeneteket egy n sort tartalmazó szöveges fájlból kapja, ahol minden sorban egy szimbólum megjelenési valószínűsége található.

1. Természetes nyelvek entrópiája

2. Entrópia, mutális információ, relatív entrópia

3. Huffman- és Shannon–Fano kódok I.

4. Huffman- és Shannon–Fano kódok II.

5. Szótár alapú kódolások

6. Hibajavító kódok

7. ISBN

1. példa:

[0.17, 0.15, 0.39, 0.04, 0.09, 0.16]

=>

2.2963538019348038

2.35

2.41

2. példa:

[0.2, 0.3, 0.4, 0.1]

=>

1.8464393446710154

1.9

1.9

1. Természetes nyelvek entrópiája

2. Entrópia, mutuális információ, relatív entrópia

3. Huffman- és Shannon-Fano kódok I.

4. Huffman- és Shannon-Fano kódok II.

5. Szótár alapú kódolások

6. Hibajavító kódok

7. ISBN

4. Huffman- és Shannon–Fano kódok II.

Feladat

Bizonyítsuk be program segítségével, hogy 3, illetve 4 forrásszimbólum esetén (a kódábécé bináris) a Huffman és a Shannon–Fano kódok mindig megegyeznek (vagyis az átlagos kódszóhosszok egyenlőek).

Végezzünk legalább 10^4 tesztet, azaz legyen

$$P(E|f_H(X)| \neq E|f_{SF}(X)|) < 10^{-4}.$$

Az átlagos kódszóhosszokat 6 tizedes pontossággal határozzuk meg.

1. Természetes nyelvek entrópiája

2. Entrópia, mutuais információ, relatív entrópia

3. Huffman- és Shannon–Fano kódok I.

4. Huffman- és Shannon–Fano kódok II.

5. Szótár alapú kódolások

6. Hibajavító kódok

7. ISBN

5. Szótár alapú kódolások

Feladat

Implementáljuk a következő algoritmusokat:

- (a) LZ77
- (b) LZ78
- (c) LZW

A program be- és kitömörítést is meg kell valósítson.

Bemenet: tetszőleges állomány / *kimenet:* a generált kódok tetszőleges formában (pl. bináris formátum, szöveges fájl) – és fordítva.

1. Természetes nyelvek entrópiája

2. Entrópia, mutuális információ, relatív entrópia

3. Huffman- és Shannon-Fano kódok I.

4. Huffman- és Shannon-Fano kódok II.

5. Szótár alapú kódolások

6. Hibajavító kódok

7. ISBN

6. Hibajavító kódok

Feladat

Adott egy bináris hibajavító kód, melyet egy bemeneti fájlból olvasunk be, minden sor egy kódszót tartalmaz. Írjunk programot, amely megmondja, hogy a kód hány hibát képes észlelni (jelezni), illetve javítani.

1. Természetes nyelvek entrópiája

2. Entrópia, mutuális információ, relatív entrópia

3. Huffman- és Shannon-Fano kódok I.

4. Huffman- és Shannon-Fano kódok II.

5. Szótár alapú kódolások

6. Hibajavító kódok

7. ISBN

1. példa:

000

111

=>

Hibák észlelése: 2

Hibák javítása: 1

2. példa:

000

011

101

110

=>

Hibák észlelése: 1

Hibák javítása: 0

1. Természetes nyelvek
entrópiája

2. Entrópia, mutuális
információ, relatív
entrópia

3. Huffman- és
Shannon-Fano kódok
I.

4. Huffman- és
Shannon-Fano kódok
II.

5. Szótár alapú
kódolások

6. Hibajavító kódok

7. ISBN

7. ISBN

Feladat

Írjunk programot, amely tetszőleges ISBN-10 és ISBN-13 azonosítót leellenőriz, azaz megmondja róla, hogy helyes-e. Ha nem helyes, akkor írjuk ki a helyes ellenőrző számjegyet (10., illetve 13. szimbólum).

A program szöveges fájlból olvassa a bemenetet, egy sorban egy db. ISBN szerepel, melyben lehetnek szóköz, illetve kötőjel karakterek.

1. Természetes nyelvek entrópiája

2. Entrópia, mutuális információ, relatív entrópia

3. Huffman- és Shannon-Fano kódok I.

4. Huffman- és Shannon-Fano kódok II.

5. Szótár alapú kódolások

6. Hibajavító kódok

7. ISBN

1. példa:

99921-58-10-7

=>

Helyes!

2. példa:

0 8044 2957 1

=>

Helytelen. Helyesen: X

3. példa:

0 - 1343 - 2147 - 1

=>

Helytelen. Helyesen: 2

1. Természetes nyelvek
entrópiája

2. Entrópia, mutuális
információ, relatív
entrópia

3. Huffman- és
Shannon-Fano kódok
I.

4. Huffman- és
Shannon-Fano kódok
II.

5. Szótár alapú
kódolások

6. Hibajavító kódok

7. ISBN