08:15

0.-1-105

Információkezelés 4. 1. EA

2009.02.11.

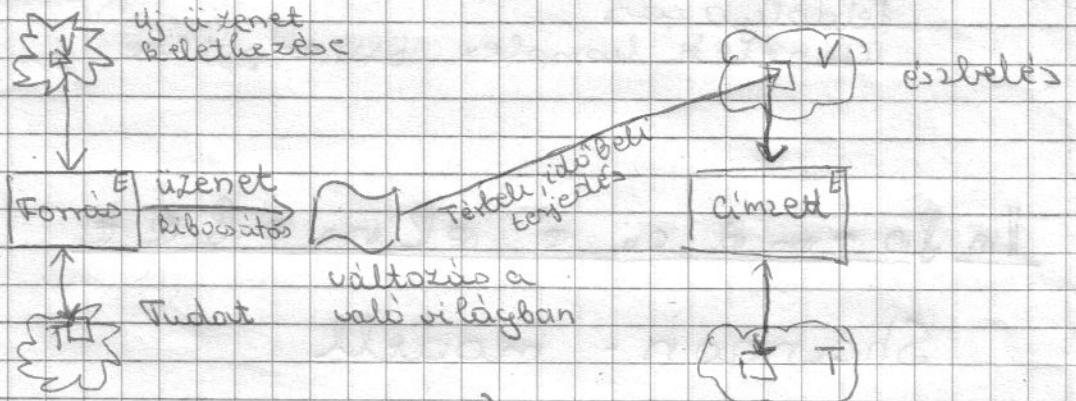
Győrfy : Információ és kódelmélet (Typotex)
Elementary Information Theory

Információelmélet alapjai

- Kommunikáció - információs rendszerek világa (véletlen jelenségekre épít)
- Shannon-féle entrópia (jövőre vonatkozik)
- Kolmogorov entrópia (számításelméletre épül) (múltra vonatkozik)

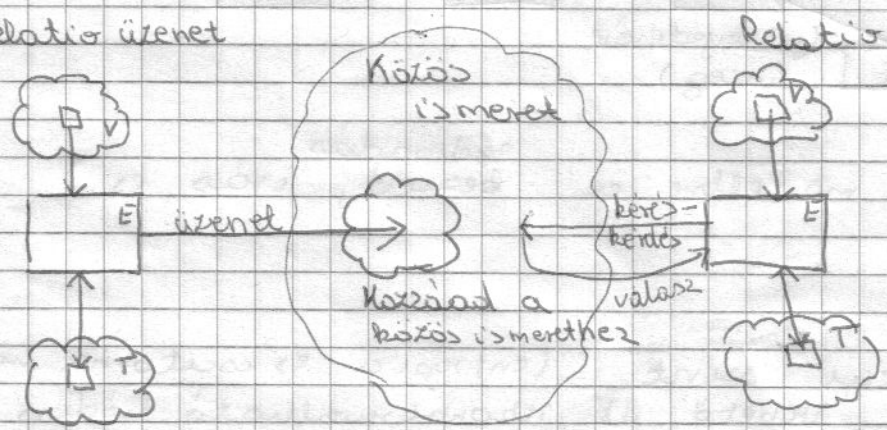
Kommunikáció lényege az emberi elmék kölcsönhatása (tudatok)

Elemi kommunikációs lépés: (információ átadás)



(Valóság - Egyén - Tudat)

Információs-rendszer jellegű kommunikáció:



Történeti állomások:

- Beszéd kialakulása
  (hangképzés (megfelelő finomságú), hallás, tudat, agyrendszer,
  (gesztikuláció - testbeszéd)

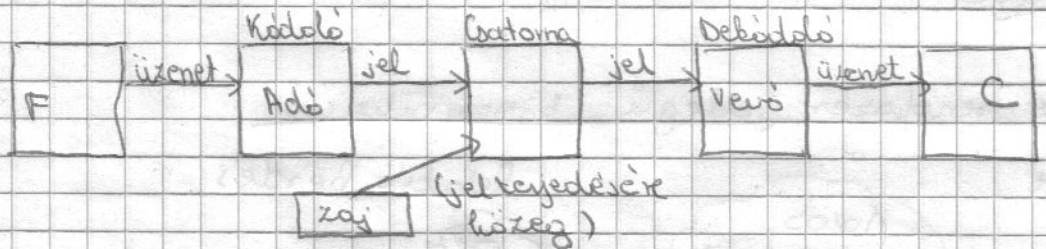
- Feljegyzés - írás, rajz, kép
  (írás megmarad)
  -- nyomtatás (sokszorosítás könnyen)
  (könyvtár ≙ szolgáltatás a leírás-kérdés elve)

- jelenleg: előbb analóg →
  - írógép
  - telefon-adatátvitel
  - mozgókép
  - számítógép
  - lemezkép
  - rádiófrekvenciás műsorszórás
  - kábelözés
  - digitális kamerák, műszerek

  → háttértár, tárolás
  számítás - mit lépés kezelni
  feldolgozás
  üzenetek komplex összeépítése

Információ elmélet

Shannon - modell



(Rákérdezhető a modellre a rádióadás, beszéd, írás is...)

Elvárás:

A) Mennyiségi szint (entrópia és csatorna kapacitás)
   -mennyire vihető át rekonstruálható módon a kívánt üzenet a csatornán?

B) Megértési szint (jelentés, szemantika)

c) Hatékonysági szint (a kívánt hatás elérése)

Mennyiségi szint:

- Forrás jellemzése: mennyi üzenetet bocsát ki, időegységre kell visszavezetni
  sebesség paraméter is kell
- Csatorna: idő egység alatt átvihető különböző csatornajelek száma

Forrásra egyszerű példa:

elemi üzenet: [igen
nem

① pénzfeldobás: fej → igen
írás → nem

Átvitel 0-1 sorozat segítségével

Feladat: 1024 × ismételjük meg

1024 0-1 bit

2^{1024} különböző jel kell hogy átmenjen a csatornán

② ötösöm van a lotón - igen
nincs - nem

1024 lét eredményét

- Ha nincs ötös - 0
- Ha 1 ötös van - 10 (sorszám / 10 bit)
- Ha több ötös van - 11 (1024 / teljes)

$P_0 = 0,999977$
$P_1 = 2,33E-5$
$P_2 = \dots$

H.F. Kódhossz várható értéke?

(Poisson határeloszlás)

Kül ① - ② - Mennyire bizonytalan az esemény kimenetelének!

3/1

2009. 02. 18., 2. előadás

ablinux.inf.elte.hu/infobe24 - segédanyagok

Mennyiségek

1. Csatorna: Kapacitás

a) bináris szimmetrikus csatornával való ekvivalencia alapján
- egységnyi idő alatt átlagosan ugyanannyi különböző jelsorozat vihető át, mint egy C bit/sec sebességű bináris csatornán

2^C különböző bit vihető át

2. Forrás: véges sok különböző szimbólumból választ (n - a szimbólumok száma)
választás - véletlen körülmények közt történik

A választást valószínűségeloszlással jellemezzük

N hosszú üzenetet választ

Eloszlás: az N hosszú üzenetek halmazán kell megadni

n^N elemű eloszlás
Shannon-entrópiáját rendeljük hozzá

- n^N elemű halmazon, elemeit sorszámozzuk $i = 1, \dots, n^N$
választási valószínűség: P_i

Shannon-entrópia:
$$H_N = -\sum_{i=1}^{n^N} P_i \log_2 P_i$$

Forrás jellemzésére H_N átlagát keressük szimbólumonként:

Def:
$$H := \lim_{N \to \infty} \frac{1}{N} H_N$$

Mérőszám: [bit/szimbólum]

Megválasztható: a forrás sebessége = sec.-ként hány szimbólumot választ: $V = \left[\frac{\text{szimbólum}}{\text{sec}}\right]$

C: Bit/sec
H: Bit/szimbólum
V: szimbólum/sec

Mennyiségi feladat megoldhatóságának: veszteségmentesen átvihető minden üzenet

Zajmentes csatorna alaptétele (Shannon)

(T) Nem lehet veszteségmentesen működtetni a kommunikációs rendszert, ha $\boxed{V > \frac{C}{H}}$;

Tetszőleges ($\varepsilon > 0$)-ra ∃ kódolás, hogy $\boxed{V < \frac{C}{H} - \varepsilon}$ sebességgel veszteségmentesen működtethető a rendszer.

Def: Csatorna kapacitás:

Diszkrét, zajmentes csatorna

Átvihető jelek: $s_1, s_2, \dots, s_m$
Átviteli idők: $t_1, t_2, \dots, t_n$

Szim. bin. csatorna: $m = 2$
$t_1 = t_2 = t$

Ha T ideig működtetjük, akkor hány különböző jelsorozatot küldhetek át?

T/t jel vihető át

$N(t) = 2^{T/t}$ -összes lehetőség

$C = 1/t$

$N(t) = 2^{C \cdot T}$

$\boxed{C = \frac{1}{T} \cdot \log_2 N(T)}$

Két csatorna ekvivalens, ha az $N(T)$ mennyiségek „lényegében" megegyeznek.

Def: Legyen a csatornán T idő alatt átvihető különböző jelsorozatok száma $N(T)$, ekkor a csatorna kapacitása: $\boxed{C = \lim_{T \to \infty} \frac{\log_2 N(T)}{T}}$

$2^{T \cdot (C - \delta)} < N(T) < 2^{T \cdot (C + \delta)}$ δ: tetszőleges kicsi: $\delta > 0$

$s_1, \dots, s_n$ / $t_1, \dots, t_m$ | ha $t_1 = t_2 = \dots = t_m = t$, akkor szimmetrikus

C megadása?

Általános eset:

$$N(T) = N(T-t_1) + N(T-t_2) + \dots + N(T-t_m)$$

(differencia egyenlet)

Megoldása:

$$1 - X^{-t_1} - X^{-t_2} - \dots - X^{-t_m} = 0$$

X_0: a legnagyobb pozitív gyök

$$X_0^T = \underbrace{(X_0^{-t_1} + X_0^{-t_2} + \dots + X_0^{-t_m})}_{1} \cdot X_0^T =$$

$$= X_0^{T-t_1} + X_0^{T-t_2} + \dots + X_0^{T-t_m}$$

[X_0 határánya úgy viselkedik mint $N(T)$

$$N(T) = k \cdot X_0^T$$

$$C = \lim_{T \to \infty} \frac{\log_2 N(t)}{T} = \boxed{\log_2 X_0 = C}$$

// T lineáris növelése exponenciálisan növeli az $N(T)$-t

Forrás jellemzése:

Shannon-entrópia levezetése

- minden lehetséges üzenetvalószínűségre működjön
- nem a szimbólumok fontosak, hanem a választás bizonytalansága (mélyen elosztás szerint?)

Bizonytalanság mérőszáma: eloszlást jellemez

Valószínűségi eloszlás: $(p_1, \dots, p_n)$, akol $\forall i: p_i > 0$ és $\sum_{i=1}^{n} p_i = 1$

Keresünk a (véges, diszkrét) eloszlásokon egy $H(P_1, \dots, P_n)$ fv-t, melyre Shannon-elvárásai:

(0) H az eloszlásokon van értelmezve, független a $\{P_1, \dots, P_n\}$ permutációitól

(1) H legyen folytonos minden változójában

(akkor is, ha $P_n \to 0$ $H(P_1, \dots, P_n) \to H(P_1, \dots, P_{n-1})$)

(2) Az $A(n) = H\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$ egyenletes eloszlás bizonytalansága

$A(n) < A(m)$, ha $n < m$

// Monotonitási elvárás a szimmetrikus szituációra

(3) Elágazási (lépcsőzési szabály)

$$H(P_1, P_2, \dots, \underbrace{P_{n-1}, P_n}_{q = P_{n-1} + P_n}) = H(P_1, P_2, \dots, P_{n-2}, q) + q \cdot H\left(\frac{P_{n-1}}{q}, \frac{P_n}{q}\right)$$

$$\boxed{H(P_1, \dots, P_n) = H(P_1, \dots, P_{n-2}, q) + q \cdot H\left(\frac{P_{n-1}}{q}, \frac{P_n}{q}\right)}$$

// Általános elágazási szabály: (indukcióval)

Legyen: $q_j = \sum_{i=1}^{n_j} P_{j,i}$ $\quad j = 1, \dots, m$

$$0 < P_{j,i} < 1 \quad , \quad \sum_{j=1}^{m} q_j = 1$$

$$H(\underbrace{P_{1,1}, \dots, P_{1,n_1}}_{q_1}, \underbrace{P_{2,1}, \dots, P_{2,n_2}}_{q_2}, \dots, \underbrace{P_{m,1}, \dots, P_{m,n_m}}_{q_m}) =$$

$$= H(q_1, \dots, q_m) + \sum_{j=1}^{m} q_j \cdot H\left(\frac{P_{j,1}}{q_j}, \dots, \frac{P_{j,n_j}}{q_j}\right)$$

↑ elágazásig tartó bizonytalanság ↑ elágazás utáni bizonytalanság

Kaptunk egy fv-egyenletet H-ra

(7.)

A(0) – (3) feltételeket kielégítő f, csak

$$k \cdot \sum_{i=1}^{n} p_i \cdot \log_2 p_i \quad \text{alakú lehet}$$

k megválasztása? amit 1 bittel tudunk mérni

$$1 = H\left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$

$$k(-1) = 1 \;, \quad \boxed{k = -1}$$

$$\boxed{H(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i} \;, \text{ ami a Shannon-entrópia formulája}$$

Biz: Racionális eloszlás: +/- <-> A kifejezése segítségével

$$p_i = \frac{q_i}{m} \qquad i = 1, \dots, n$$

$$\sum_{i=1}^{n} p_i = 1 \quad \Rightarrow \quad \sum_{i=1}^{n} q_i = m$$

Az m elemű egyenletes eloszlásból indulva -csoportok képzése

$$\underbrace{H\left(\frac{q_1}{m}, \dots, \frac{q_n}{m}\right)}_{\text{elágazásig}} = \underbrace{A(m)}_{\text{teljes}} - \sum_{i=1}^{n} \frac{q_i}{m} \cdot \underbrace{A(q_i)}_{\text{elágazás utáni}} =$$

$$= -\sum_{i=1}^{n} \frac{q_i}{m} \cdot \left[A(q_i) - A(m)\right]$$

Ha $A(n) = k \cdot \log_2 n$ alapú (Be kell Biz.)

$\Rightarrow$ Következik a Shan. és kész a racionális eset és a valós az már a folytonosságból lejön

Egyenletes eloszlás entrópiája: $\boxed{A(n) = k \cdot \log_2 n}$

A(n) meghatározása:

$$A(n \cdot m) = H\left(\underbrace{\tfrac{1}{nm}, \dots, \tfrac{1}{nm}}_{m}, \dots, \underbrace{\tfrac{1}{nm}, \dots, \tfrac{1}{nm}}_{m}\right) = A(n) + n \cdot \tfrac{1}{n} \cdot A(m)$$

elág.-ig elág. után

5/2

$$\boxed{A(n \cdot m) = A(n) + A(m)}$$

$$A(n^m) = m \cdot A(n)$$

Rögzítsük az s egész számot: $s^m \leq t^n \leq s^{m+1}$

(t paraméter, egész szám)

$$s^m \leq t^n \leq s^{m+1} \quad \Big| \text{ alkalmazzuk: } \begin{matrix} \log_2 x \\ A(x) \end{matrix} \Big\} \text{monoton} \Rightarrow \leq \text{marad}$$

$$\log_2 s^m \leq \log_2 t^n \leq \log_2 s^{m+1} \quad \Big| \quad A(s^m) \leq A(t^n) \leq A(s^{m+1})$$

$$m \cdot \log_2 s \leq n \cdot \log_2 t \leq (m+1) \cdot \log_2 s \quad \Big| \quad m \cdot A(s) \leq n \cdot A(t) \leq (m+1) \cdot A(s)$$

$$\frac{m}{n} \log_2 s \leq \log_2 t \leq \frac{m+1}{n} \cdot \log_2 s \quad \Big| \quad \frac{m}{n} A(s) \leq A(t) \leq \frac{m}{n} A(s) + \underbrace{\frac{1}{n} \cdot A(s)}_{\downarrow n \to \infty \; 0}$$

$$\underbrace{\frac{m}{n} \log_2 s + \underbrace{\frac{1}{n} \log_2 s}_{\downarrow n \to \infty \; 0}}$$

$$\frac{m}{n} \leq \frac{\log_2 t}{\log_2 s} \leq \frac{m}{n} + \frac{1}{n} \quad \Big| \quad \frac{m}{n} \leq \frac{A(t)}{A(s)} \leq \frac{m}{n} + \frac{1}{n}$$

$$n \to \infty : \quad \frac{m}{n} \Rightarrow \frac{\log_2 t}{\log_2 s} \quad \Big| \quad \frac{m}{n} \longrightarrow \frac{A(t)}{A(s)}$$

$$\Rightarrow \frac{A(t)}{A(s)} = \frac{\log_2 (t)}{\log_2 (s)}$$

$$A(t) = \underbrace{\frac{A(s)}{\log_2 s}}_{R} \cdot \log_2 t$$

$$\boxed{H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \log_2 n}$$

3. előadás, 2009. 02. 25.

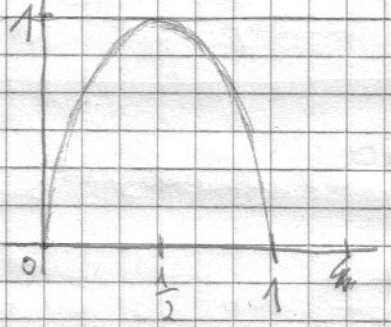
$$H(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

Tulajdonságai:

① $H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) =: A(n) = \log_2 n$

② Kételemű eloszlás:

$$H(p, 1-p) = -p \cdot \log p - (1-p) \cdot \log_2 (1-p)$$



$H(p, 1-p) \leq 1$

Ha p csökken, H csökken

③ $H(p_1, \dots, p_n) \leq \log_2 n$ (egyenletességi növekszik)

$$H(p_1, \dots, \underbrace{p_{n-1} + p_n}) = H(p_1, \dots, p_{n-2}, p) + p \cdot H\left(\frac{p_{n-1}}{p}, \frac{p_n}{p}\right)$$

$$H(p_1, \dots, p_{n-2}, \underbrace{q_1, q_2}) = H(p_1, \dots, p_{n-2}, p) + p \cdot H\left(\frac{q_1}{p}, \frac{q_2}{p}\right)$$

$(p_{n-1} + p_n = q_1 + q_2 = p)$

⇒ Kiegyenlítés növeli az entrópiát

(Forrás amit kibocsát = val. szín. változó)

Valószínűségi változó entrópiája

ξ, értékei: $x_1, \dots, x_n$ (diszkrét, véges)

eloszlása: $P(\xi = x_i)$ $i = 1, \dots, n$

$$\sum_{i=1}^{n} P(\xi = x_i) = 1$$

$$\boxed{H(\xi) := -\sum_{i=1}^{n} P(\xi = x_i) \cdot \log_2 P(\xi = x_i)}$$

①/3 Forrás jellemzése: N db v. sz. változóval

Kétdimenziós eloszlás entrópiája: ξ, η

Együttes eloszlás: $P(\xi = x_i, \eta = y_j)$ $\quad i = 1 \dots n$, $j = 1 \dots m$

$$H(\xi, \eta) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(\xi = x_i, \eta = y_j) \cdot \log_2 P(\xi = x_i, \eta = y_j)$$

ξ eloszlása?

$$P(\xi = x_i) = \sum_{j=1}^{m} P(\xi = x_i, \eta = y_j)$$

$$P(\eta = y_j) = \sum_{i=1}^{n} P(\xi = x_i, \eta = y_j)$$

Független eloszlások entrópiája:

$$H(\xi, \eta) ; \; H(\xi) \; , \; H(\eta)$$

ξ, η függetlenek, akkor $P(\xi = x_i, \eta = y_j) = P(\xi = x_i) \cdot P(\eta = y_j$

$$H(\xi, \eta) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(\xi = x_i) \cdot P(\eta = y_j) \cdot [\log_2 P(\xi = x_i) + \log_2 P(\eta = y_j)] =$$

$$= -\sum_{i=1}^{n} \sum_{j=1}^{m} P(\xi = x_i) \cdot P(\eta = y_j) \cdot \log_2 P(\xi = x_i) +$$

$$+ -\sum_{i=1}^{n} \sum_{j=1}^{m} P(\xi = x_i) \cdot P(\eta = y_j) \cdot \log_2 P(\eta = y_j) =$$

$$= -\sum_{i=1}^{n} P(\xi = x_i) \cdot \log_2 P(\xi = x_i) \cdot \underbrace{\sum_{j=1}^{m} P(\eta = y_j)}_{=1}$$

$$+ \underbrace{\sum_{i=1}^{n} P(\xi = x_i)}_{=1} \cdot \sum_{j=1}^{m} \left(\log_2 P(\eta = y_j)\right) P(\eta = y_j)$$

$$= H(\xi) + H(\eta)$$

$$\boxed{H(\xi, \eta) = H(\xi) + H(\eta)}$$

(Várhatóan ez a legnagyobb bizonytalanság)

Tétel: $H(\xi, \eta) \leq H(\xi) + H(\eta)$ és $= \Leftrightarrow$ a.cs.a. ξ, η függetlenek

Biz:

Tegyük fel, h. megfigyeltük $\eta = y_j$ értéket

$P(\xi = x_i \mid \eta = y_j) \quad i = 1, \dots, n$

$P(A|B) = \frac{P(AB)}{P(B)}$

$$P(\xi = x_i \mid \eta = y_j) = \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)}$$

$$H(\xi \mid \eta = y_j) = -\sum_{i=1}^{n} P(\xi = x_i \mid \eta = y_j) \cdot \log_2 P(\xi = x_i \mid \eta = y_j) \neq$$

$\neq$

Feltételes entrópia:

Mennyi bizonytalanság marad ξ-re nézve, ha η-t megfigyeltem?

$H(\xi, \eta)$ - ha egyiket sem figyeltem meg belőle

ebből eltűnik η megfigyelésének bizonytalansága: $\boxed{H(\xi, \eta) - H(\eta) =: H(\xi \mid \eta)}$

ξ feltételes entrópiája η szerint

Elágazási szabályt felhasználva:

$H(\eta)$ - elágazásig tartó

$H(B|\eta)$ - elágazás utáni bizonytalanság

$$H(\xi \mid \eta) = -\sum_{j=1}^{n} P(\eta = y_j) \cdot \sum_{i=1}^{n} P(\xi = x_i \mid \eta = y_j) \cdot \log_2 P(\xi = x_i \mid \eta = y_j)$$

$$= -\sum_{j=1}^{n} P(\eta = y_j) \cdot \sum_{i=1}^{n} \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} \cdot \log_2 \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)}$$

$$= \boxed{\sum_{j=1}^{n} \sum_{i=1}^{n} P(\xi = x_i, \eta = y_j) \cdot \log_2 \frac{P(\eta = y_j)}{P(\xi = x_i, \eta = y_j)}}$$

$$H(\xi,\eta) \le H(\xi) + H(\eta)$$

$$H(\xi,\eta) - H(\eta) \le H(\xi)$$

Tétel (másik alakja)

$$H(\xi|\eta) \le H(\xi)$$ a cs.a. =, ha függetlenek

T.h. megfigyelhetjük η-t, vagy η egy $f(\eta)$ függvényét.

Lemma

$$H(\xi|\eta) \le H(\xi|f(\eta))$$ (Kevesebb ismeret az entrópiát kevésbé csökkenti)

és egyenlőség feltétele:

$\forall z$-re, amire $f(y)=z$

$$P(\xi=x|\eta=y) = P(\xi=x|f(\eta)=z)$$

—

$(\xi, f(\eta))$

$$P(\xi=x, f(\eta)=z) = \sum_{f(y)=z} P(\xi=x, \eta=y)$$

$$P(f(\eta)=z) = \sum_{f(y)=z} P(\eta=y)$$

$$H(\xi|f(\eta)) = \sum_x \sum_z \left[\sum_{f(y)=z} P(\xi=x, \eta=y)\right] \cdot \log_2 \frac{\sum_{f(y)=z} P(\eta=y)}{\sum_{f(y)=z} P(\xi=x, \eta=y)}$$

$$H(\xi|\eta) = \sum_x \underbrace{\sum_z \sum_{f(y)=z}}_{=\sum_y} P(\xi=x, \eta=y) \cdot \log_2 \frac{P(\eta=y)}{P(\xi=x, \eta=y)}$$

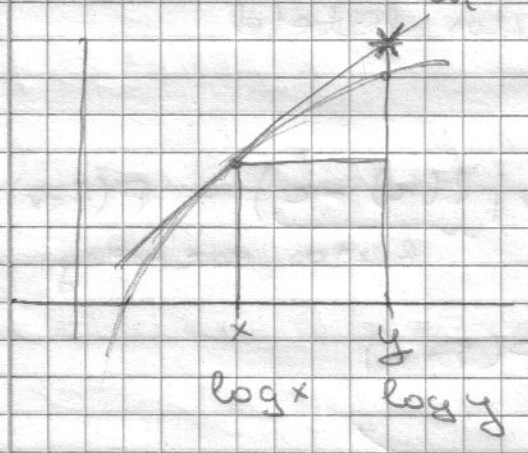
csak a $\sum_x \sum_z$ -n belüli összegek kell valószínűségi elek típusa: a_i $i=1,\dots,n$ $a=\sum a_i$

b_j $j=1,\dots,n$ $b=\sum b_j$

$$\sum_{i=1}^{n} a_i \cdot \log \frac{b_i}{a_i} \qquad (H(\xi|\kappa))$$

$$\leq a \cdot \log \frac{b}{a} \qquad (H(\xi|f(\kappa)))$$

= feltétele: $\frac{b_i}{a_i} = \frac{b}{a}$ $\quad i = 1 \dots n$



* lilából becsüli $\log y$-t

$$\log y \leq \log x + c_x (y - x)$$
$$= a \cdot \text{és } a \quad x = \frac{b}{a}$$

[lényeg, h. lehúzzuk $\frac{b}{a}$-t ($\log_2$)]

$$\log \frac{b_i}{a_i} \leq \log \frac{b}{a} + c\left(\frac{b_i}{a_i} - \frac{b}{a}\right)$$

$= a \cdot \text{és } a \quad \frac{b_i}{a_i} = \frac{b}{a}$ $\quad \forall i$-re

$$* a_i \Rightarrow \sum_{i=1}^{n} a_i \log \frac{b_i}{a_i} \leq \sum_{i=1}^{n} a_i \cdot \log \frac{b}{a} + c \sum_{i=1}^{n} \left(\frac{b_i}{a_i} - \frac{b}{a}\right)$$

$$\dots \leq \log \frac{b}{a} \cdot \sum_{i=1}^{n} a_i + \dots$$

$$\dots \leq a \cdot \log \frac{b}{a} + \dots$$

$$\vdots$$

$$z\text{-re: } \sum_{f(y)=z} P(\xi = x, \kappa = y) \log_2 \frac{P(\kappa = y)}{P(\xi = x, \kappa = y)} \leq$$

$$\leq \left(\sum_{f(y)=z} P(\xi = x, \kappa = y)\right) \cdot \log_2 \frac{\sum_{f(y)=z} P(\kappa = y)}{\sum_{f(y)=z} P(\xi = x, \kappa = y)}$$

$$\Leftrightarrow = : \frac{P(\xi = x, \eta = y)}{P(\kappa = y)} = \frac{\sum_{f(y)=z} P(\xi = x, \kappa = y)}{\sum_{f(y)=z} P(\kappa = y)}, \text{ ha } f(y) = z$$

$$P(\xi = x \mid \kappa = y) = P(\xi = x \mid f(\kappa) = z)$$

■ lemma

lemmából, hogyan bizonyítsuk be, hogy:

$$H(\xi|\eta) \leq H(\xi)$$

válasszuk az $f(y) \equiv c$ fv.t.

$\Rightarrow H(\xi|\eta) \leq H(\xi|f(\eta)) = H(\xi)$
↑ biztos feltétel

= feltétele: $f(y) = c$

$$P(\xi = x | \eta = y) = P(\xi = x | \underbrace{f(\eta) = c}_{\text{biztos esemény}}) = P(\xi = x)$$

$$\frac{P(\xi = x, \eta = y)}{P(\eta = y)} = P(\xi = x)$$

$\Downarrow$ függetlenek

múlt ← jövő
ξ_1, ξ_2, ξ_3
jelen

lemma bizonyítása: ▨ tétel

$$H(\xi_3|\xi_2, \xi_1) \leq H(\xi_3|\xi_2)$$

$=$ akkor és csak akkor, ha $P(\xi_3 = x | \xi_1 = y, \xi_2 = z) = P(\xi_3 = x | \xi_2 = z)$

"a legnagyobb bizonytalanság a jövőre akkor van, ha csak a jelent ismerem és a múltat nem"

(Markovlánc tulajdonság: "a jelen ismeretében a múlt és a jövő független"

"A jövő a múltól csak a jelenen keresztül függ"

2009.03.04., 4. előadás

Kiült órán: $\sum a_i \log \frac{b_i}{a_i} \leq a \log \frac{b}{a}$

ahol $a = \sum a_i$

$b = \sum b_i$

Következmények:

① $(p_1, \dots, p_n)$ $(q_1, \dots, q_n)$ $\sum 1$ eloszlások

a_i b_j

$$\sum_{i=1}^{n} p_i \log_2 \frac{p_i}{q_i} \leq 0$$

(ui. $\log_{\frac{1}{2}} \frac{1}{1} = 0$)

$$-\sum_{i=1}^{n} p_i \log_2 p_i \leq -\sum_{i=1}^{n} p_i \log_2 q_i$$

$\uparrow 1$

② $b_i = 1 \quad i = 1, \dots, n$

$a_i \leftarrow (p_1, \dots, p_n)$

$$\sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \leq \log_2 \frac{n}{1} = \log_2 n$$

(egyenletes entrópia maximális)

③ Jensen-egyenlőtlenség

$h(x)$ konvex fv, ξ - valós. változós

↙ várható érték

$E(h(\xi)) \geq h(E(\xi))$

(esetünkben h a logaritmus, E az entrópia)

Forrás jellemzése:

Tipikus közelítések: a lehető legbizonytalanabb nálra készülünk, azaz felülről kell közelíteni a forrás entrópiáját

Választható szimbólumok: $x_1, \dots, x_n$

0-ad rendű közelítés - nincs megfigyelés

- Egyenletes eloszlást tételezünk fel, ξ

ξ eloszlása $P(\xi = x_i) = \frac{1}{n} \quad \forall i = 1, \dots, n$

- függetlenek a választások, N hosszú választásnál $(\xi_1, \dots, \xi_N)$

$$P(\xi_1 = x_{i_1}, \dots, \xi_N = x_{i_N}) = P(\xi_1 = x_{i_1}) \cdot \dots \cdot P(\xi_N = x_{i_N}) = \left(\frac{1}{n}\right)^N$$

$$H = \lim_{N \to \infty} \frac{1}{N} \cdot H(\xi_1, \dots, \xi_N)$$

↗ egy szimbólumra jutó átlagos entrópia

$$H(\xi_1, \dots, \xi_n) = N \log_2 n \quad \text{(egyenletes)}$$

$$H = \lim_{N \to \infty} \frac{1}{N} N \log_2 n \; , \quad \boxed{H = \log_2 n}$$

1-rendű közelítés:

M hosszú üzenetet megfigyelünk

x_i előfordulása: $m_i \quad \forall i = 1, \dots, n$

Eloszlás közelítése $P_i := P(\xi = x_i) = \frac{m_i}{M}$, $i = 1, \dots, n$

$$P(\xi_1 = x_{i_1}, \dots, \xi_N = x_{i_N}) = \prod_{j=1}^{N} P(\xi_j = x_{i_j}) = \prod_{i=1}^{n} P_i^{M_i}$$

(akkor a kifejtés legbizonytalanabb, ha függetlenek)

M_i - x_i szimbólum hány szor fordul elő a produktumban, azaz P_i hányszor fordul elő

$$H(\xi_1, \dots, \xi_N) = \sum_{i=1}^{n} H(\xi_i) = N H(\xi)$$

$$H = -\sum_{i=1}^{n} P(\xi = x_i) \cdot \log_2 P(\xi = x_i)$$

2-ordrendű közelítés:

M megfigyelés, m_i - x_i előfordulásainak száma

m_{ij} - az $x_i x_j$ egymás utáni előfordulásainak a száma

$P(\xi = x_i) = \frac{m_i}{M}$ - önmagában való választás

$P(\xi_t = x_i, \xi_{t+1} = x_j) = \frac{m_{ij}}{M}$ - egymás utáni párokra való vetületi (peremeloszlás)

$$p_{j|i} := P(\xi_{t+1} = x_j \mid \xi_t = x_i) = \frac{P(\xi_t = x_i,\ \xi_{t+1} = x_j)}{P(\xi_t = x_i)} = \frac{\frac{m_{ij}}{M}}{\frac{m_i}{M}} = \frac{m_{ij}}{m_i}$$

$$p_{j|i} = \frac{m_{ij}}{m_i}$$

$$P(\xi_1 = x_{i_1}, \dots, \xi_N = x_{i_N}) = P(\xi_N = x_{i_N} \mid \xi_1 = x_{i_1}, \dots, \xi_{N-1} = x_{i_{N-1}}) \cdot$$
$$\cdot P(\xi_1 = x_{i_1}, \dots, \xi_{N-1} = x_{i_{N-1}})$$

// Markovlánc tulajd.: $H(\xi_3 \mid \xi_1, \xi_2) \leq H(\xi_3 \mid \xi_2)$

Markov-lánc: (ekkor van a legnagyobb bizonyt.)

$$P(\xi_N = x_{i_N} \mid \xi_1 = x_{i_1}, \dots, \xi_{N-1} = x_{i_{N-1}}) = P(\xi_N = x_{i_N} \mid \xi_{N-1} = x_{i_{N-1}})$$

Teljesüljön $\forall x_{i_N}, x_{i_{N-1}}$ értékre N-től függetlenül

$$= p_{i_N | i_{N-1}}$$

homogén, egylépéses Markovlánc:

$$P(\xi_{t+1} = x_i \mid \xi_1 = x_{i_1}, \dots, \xi_{t-1} = x_{i_{t-1}}, \xi_t = x_j) =$$
$$= P(\xi_{t+1} = x_i \mid \xi_t = x_j) = p_{i|j}$$

$p_{i|j}$ = átmeneti valószínűségek

(átmenet valószínűség mátrix, Σ egysor $= 1$)

$$P(\xi_1 = x_{i_1}, \dots, \xi_N = x_{i_N}) = P(\xi_1 = x_{i_1}, \dots, \xi_{N-1}) \cdot P(\xi_N = x_{i_N} \mid \dots)$$
$$= \dots = P(\xi_1 = x_{i_1}) \cdot P(\xi_2 = x_{i_2} \mid \xi_1 = x_{i_1}) \cdot \dots \cdot P(\xi_N = x_{i_N} \mid \xi_{N-1} = x_{i_{N-1}})$$

M_{ij}: $P(\xi_t = x_i \mid \xi_{t+1} = x_j)$ alakú tényezők száma

$$P(\xi_1 = x_{i_1}, \dots, \xi_N = x_{i_N}) = P(\xi_1 = x_{i_1}) \cdot \prod_{i,j} p_{i|j}^{M_{ij}}$$

$$H(\xi_1, \dots, \xi_N) = H(\xi_1, \dots, \xi_{N-1}) + \underbrace{H(\xi_N \mid \xi_1, \dots, \xi_{N-1})}_{\text{akkor max, ha csak az utolsótól függ}} =$$

$$= H(\xi_1, \dots, \xi_{N-1}) + H(\xi_N \mid \xi_{N-1}) = \dots =$$
$$= H(\xi_1) + H(\xi_2 \mid \xi_1) + \dots + H(\xi_N \mid \xi_{N-1})$$

$$H(\xi_t \mid \xi_{t-1}) = \sum_{i=1}^{n} P(\xi_{t-1} = x_i) \cdot H(\xi_t \mid \xi_{t-1} = x_i)$$

$$H(\xi_t \mid \xi_{t-1} = x_i) = -\sum_{j=1}^{n} P(\xi_t = j \mid \xi_{t-1} = i) \cdot \log_2 P_{j|i}$$

$$= -\sum_{j=1}^{n} P_{j|i} \cdot \log_2 P_{j|i} = H(\xi \mid i)$$

$$H(\xi_t \mid \xi_{t-1}) = \sum_{i=1}^{n} P(\xi_{t-1} = x_i) \cdot H(\xi \mid i)$$

$P(\xi_t = x_i)$

$t \to \infty$

Markov-lánc .... ergodikus tulajdonság

$$\lim_{t \to \infty} P(\xi_t = x_i) = \Pi_i$$

(Feltétel: minden érték visszatérő legyen)
ne legyen periodikus

Markov folyamat határeloszlása - stacionárius eloszl.

$$P(\xi_t = x_i) = P(\xi_{t-1} = x_j) \cdot P_{i|j} \qquad \forall\ i\text{-re}$$

$$\lim_{t \to \infty} \Pi_i = \sum_{j=1}^{n} \Pi_j \cdot P_{i|j} \qquad \forall\ i\text{-re}$$

(ez egy lin. egy. rendszer)

$(\Pi_1, \dots, \Pi_n)$ az átmenetvalószínűségi mátrix sajátvektora a $\lambda = 1$ sajátértékkel.

Legyen: $P(\xi_t = x_i) = \Pi_i$ - ettől kezdve $P(\xi_t = x_i) = \Pi_i$ marad, az időtől nem függ (azaz stacion.)

$$H(\xi_1, \dots, \xi_N) = H(\xi_1) + \sum_{j=1}^{n} H(\xi \mid j) \cdot \left(\sum_{t=1}^{N} P(\xi_{t-1} = x_j) \right)$$

$$\lim_{N \to \infty} \frac{1}{N} \cdot H(\xi_1, \dots, \xi_N) = \sum_{j=1}^{n} H(\xi \mid j) \cdot \left(\lim_{N \to \infty} \frac{1}{N} \sum_{t=2}^{N} P(\xi_{t-1} = x_j) \right)$$

$\downarrow \Pi_j$

$$= \sum_{j=1}^{n} \Pi_j \cdot H(\xi \mid j)$$

$$H(\xi_1, \ldots, \xi_N) = \sum_{j=1}^{n} \pi_j \cdot H(S|_j)$$

3-adrendű:

m_{ijk} – $x_i x_j x_k$ egymás után

$P(\xi_t = x_k \mid \xi_{t-1} = x_j, \xi_{t-2} = x_i)$ – rögzítjük $\frac{m_{ijk}}{m_{ij}}$

Kétlépéses Markov lánc:

$P(\xi_t = x_k \mid \xi_1 = x_{i_1}, \ldots, \xi_{t-3} = x_{i_{t-3}}, \xi_{t-2} = x_{i_k} \mid \xi_{t-1} = x_j)$

$= P(\xi_t = x_k \mid \xi_{t-2} = x_i, \xi_{t-1} = x_j)$

(betűszimbólumból állítunk a szavak megfigyelésére ...)

(És a levezetés nem kell vizsgára)

Infók 4., 5. előadás, 2009. 03. 11.

Hosszú sorozatok jellemzése:

"nagy számok törvénye" típusú

Forrás: független, azonos eloszlású // legrosszabb eset // legbizonytalanabb

$X_1, \dots, X_n$ - választható jelek

$P(\xi = X_i) = P_i$

$P(\xi_1 = X_{i_1}, \dots, \xi_N = X_{i_N}) = \prod_{i=1}^{n} P_i^{M_i}$ // M_i - ez X_i előfordulások száma

$P(\vec{\xi} = \vec{x})$

(relatív gyakoriság tart a vsz.-hez) ← NSZT

$M_i = N \cdot P_i \quad , i = 1 \dots n$

$$P(\vec{\xi} = \vec{x}) = \prod_{i=1}^{n} P_i^{N \cdot P_i} = 2^{-\sum N \cdot P_i \cdot \log_2 P_i} = 2^{-N \cdot H}$$

Tétel: Tetszőleges $\varepsilon > 0$, $\delta > 0$ - hoz létezik N_0, hogy $N > N_0$ esén az N hosszú sorozatok két halmazba sorolhatók:

(1) lényegtelen halmaz, annak a vsz.-e, hogy ebből választunk $< \varepsilon$

(2) Tipikus halmaz (ebbe $1-\varepsilon$-nál nagyobb vsz.-gel esik a választás)

$\vec{x}$ tipikus

$$\left| \underbrace{\frac{\log_2 \frac{1}{P(\vec{\xi} = \vec{x})}}{N}}_{\text{közelítőleg } H} - H \right| < \delta$$

// Nagyon hosszú jelsorozat esetén a tipikus halmazon majdnem egyenletes teszi a választást

Szerepe a tételnek:

Tipikus halmaz: $|M_N|$ - adjunk becslést az elemszámára

Becsüljük a $P(\vec{\xi} = \vec{x})$ értékét, ahol $\vec{x}$ tipikus:

$$2^{-N(H+\delta)} < P(\vec{\xi} = \vec{x}) < 2^{-N(H-\delta)}$$

$$|M_n| < 2^{N(H+\delta)}$$

Biz: (tétel)

$\vec{\xi}$ $\qquad \prod_{i=1}^{n} p_i^{M_i}$

$M_i \qquad i=1 \ldots n$

M_i-t előállítani: $\xi_j^{(i)} \quad j=1,\ldots,N \quad \xi_j^{(i)} = \begin{cases} 0, & \text{ha } \xi_j \neq x_i \\ 1, & \text{ha } \xi_j = x_i \end{cases}$

$$M_i = \sum_{j=1}^{N} \xi_j^{(i)}, \quad P(\xi_j^{(i)} = 1) = p_i$$

$M(M_i) = N \cdot M(\xi^{(i)}) = N \cdot p_i$ $\qquad$ M - várható érték,

$D^2(M_i) = N \cdot D^2(\xi^{(i)}) = N \cdot (p_i - p_i^2) = N \cdot p_i \cdot (1-p_i)$ $\qquad$ D - szórás $[= E(\xi^2) - E^2(x)]$

Csebisev-egyenlőtlenség:

$$P((\xi - M(\xi))^2 > \gamma^2) \leq \frac{D^2(\xi)}{\gamma^2}$$

Alkalm.

$$P((M_i - Np_i)^2 > (\delta' N)^2) \leq \frac{N \cdot p_i \cdot (1-p_i)}{(\delta' N)^2} < \frac{\text{Konstans}}{N} < \varepsilon'$$

Megválasztjuk N-et olyan nagynak, hogy ε'-től függően ez teljesüljön $\forall\, i = 1 \ldots n$-re.

Erre az N-re

$$P(\text{legalább egy } i\text{-re } |M_i - N \cdot p_i| > \delta' N) \leq n \cdot \varepsilon'$$

$\mathcal{E}$ - lényegtelen halmaz: $\varepsilon' = \frac{\varepsilon}{n}$

A komplementere: tipikus halmaz:

$$\forall\, i\text{-re } |M_i - Np_i| < \delta' N$$

$$N(p_i - \delta') < M_i < N(p_i + \delta')$$

$$\prod_{i=1}^{n} p_i^{N(p_i+\delta')} < \prod_{i=1}^{n} p_i^{M_i} < \prod_{i=1}^{n} p_i^{N(p_i-\delta')} \qquad // \; p_i < 1$$

↖ $\log_2$-t véve

$$N \cdot \left(\sum_{i=1}^{n} \log_2 p_i (p_i + \delta')\right) < \log_2 P(\vec{\xi} = \vec{x}) < N \cdot \left(\sum_{i=1}^{n} (p_i - \delta') \cdot \log_2 p_i\right)$$

// H behoz egy mínusz és a P^{-1} is fordít az előjelen

$$H + \delta' \sum_{i=1}^{n} \log_2 p_i < \frac{\log_2 P^{-1}(\vec{\xi} = \vec{x})}{N} < H - \underbrace{\delta' \sum_{i=1}^{n} \log_2 p_i}_{\delta}$$

(2)/5

Tipikus halmaz (tulajdonsága)

A legvalószínűbb választások megadásával:

$P_1 \geq P_2 \dots \geq P_n$

legyen $L(\lambda)$ db x_i, $M_1 = N$
és nincs benne a tipikus halmazban!

$\vec{x_1}, \vec{x_2}, \dots$

$P(\vec{\xi} = \vec{x_i}) \leq P(\vec{\xi} = \vec{x}_{i-1})$

lényeges sorozat $\Rightarrow$ $\sum_{i=1}^{L} P(\vec{\xi} = \vec{x_i}) \geq 1-\lambda$

$\sum_{i=1}^{L-1} P(\vec{\xi} = \vec{x_i}) < 1-\lambda$

$2^{N(H-\delta)} < L(\lambda) < 2^{N(H+\delta)}$

↑ tipikus halmaz elemszámának becslése igaz rá

$L(\lambda)$ és M_N nem ugyanaz, de bevés elem esik a metszetükön kívül.

Kódolás és csatornakapacitás viszonya:

Forrás: $\xi_1, \dots, \xi_N$ kibocsátott

ezt a csatornán $\eta_1, \dots, \eta_{M(\xi_1,\dots,\xi_N)}$ formában vissuk át

legegyszerűbb kódolás: $f(\xi_1, \dots, \xi_N) = (\eta_1, \dots, \eta_{M(\xi_1,\dots,\xi_N)})$

$H(f(\xi)) \leq H(\xi)$ és "=" $\Leftrightarrow$ ha f invertálható

(a közvetlen megfigyeléshez képest a közvetett megfigyelés csak veszthet információt)

$\Downarrow$

$\hookrightarrow H(\xi_1, \dots, \xi_n) \geq H(\eta_1, \dots, \eta_{M(\xi_1,\dots,\xi_N)})$

Nincs veszteség: invertálható a kódolás, azaz $\exists f^{-1}$:

$f^{-1}(\eta_1, \dots, \eta_M) = (\xi_1, \dots, \xi_N)$

$\Rightarrow H(\eta_1, \dots, \eta_N) \geq H(\xi_1, \dots, \xi_N)$

Tehát, ha nincs veszteség: $H(\eta_1, \dots, \eta_M) = H(\xi_1, \dots, \xi_N)$

Ha VDA-val kódolok, akkor a kódoló automatának még van egy plusszállapota is van: $+\lambda$

A csatorna kimenetén ezt nem észlelem, kisebb az entrópia...

—>—

T ideig működik a csatorna, T hosszú jelsorozat
Mekkora ekkor a csatorna kimenetén a maximális entrópia? C bit/sec

Különböző T hosszú jelsorozatok száma: $N(T)$

$$C = \lim_{T\to\infty} \frac{\log_2 N(T)}{T}$$

Maximális: egyenletes eloszlás esetén: $\log_2 N(T) = A$, akkor

$2^{T(C-\delta)} < N(T) < 2^{T\cdot(C+\delta)}$ $\Rightarrow T\cdot(C-\delta) < H < T\cdot(C+\delta)$

(T ideig működő forrás entrópiája nem haladhatja meg a T ideig működő csatorna max. entrópiáját)

Zajmentes csatorna alaptételének bizonyítása

Forrás: H bit/szimbólum
Csatorna: C bit/sec

Tétel:

a) Nem lehet C/H-nál nagyobb sebességgel működtetni a forrást úgy, hogy V üzenet a csatorna kimenetéből visszaállítható legyen (Nem lehet veszteségmentes)

b) Tetszőleges $\delta > 0$-hoz létezik kódolás, hogy $C/H - \delta$ sebességgel veszteségmentesen lehet üzenetet átvinni.

Biz:

a) T idő, V sebesség szimb./sec
Mennyi a forrás entrópia T idő alatt: $H\cdot T\cdot V$-nél kisebb

$$V\cdot T\cdot(H-\varsigma) < H(T) < V\cdot T(H+\varsigma)$$

Csatorna kimenet max. entrópiája: $(T(C-\delta), T\cdot(C+\delta))$ közötti

$VT\cdot(H-\varsigma) > T\cdot(C+\delta)$ - ez biztos veszteséges

$VT(H-\varsigma) \leq T\cdot(C+\delta)$

$V \leq \frac{C+\delta}{H-\varsigma}$, δ, ς tetszőleges kicsi

Tehát V tetszőleges közelébe kerülhetünk

2009.03.17., 6. előadás, Infókor 4.

$\delta > 0$ $\frac{C}{H} - \delta$ sebességgel veszteségmentes legyen az üzenetátvitel

(görög betű: tetszőleges kicsi lehet, és ha üzenethosszúság elég nagy az N, akkor az egyenlőtlenség teljesül)

N hosszú üzenet, kérdés: milyen hosszú csatornajel tartozik hozzá

(Általában igaz ergodikus forrásra)

Független, azonos eloszlású forrás:

$X_1, \dots, X_n$

N hosszú üzenetek:

(i) Tipikus üzenetek, $1-\varepsilon$-nél nagyobb valószínűséggel elemszáma $\leq 2^{N(H+\delta)}$

(ii) Nem tipikus üzenetek, ε-nál kisebb vsz-gel elemszáma: $\leq n^N = 2^{N \log_2 n}$

(i) halmaz kódolására: T_1

$$2^{T_1(C-\gamma)} \geq 2^{N(H+\delta)}$$

$$T_1 \geq N \cdot \frac{H+\delta}{C-\gamma} = N \cdot \frac{H}{C} + \gamma_1$$

(ii) Marad felhasználatlan T_1 hosszú csatornajel, vegyünk egy ilyet, után T_2 hosszú jelet kódoljuk a nem tipikus üzeneteket

$$T_2 > N \frac{\log_2 n}{C-\gamma} = N \cdot \frac{\log_2 n}{C} + \gamma_2$$

A kódhossz várható értéke:

$$(1-\varepsilon)\cdot T_1 + \varepsilon\cdot(T_1+T_2) = N\cdot\left(\frac{H}{C}+\gamma_1\right) + \varepsilon\left(N\,\frac{\log_2 n}{C}+\gamma_2\right)$$

1 szimbólumra jutó kódhossz várható értéke:

$$\frac{H}{C}+\gamma_1+\gamma_3 = \frac{H}{C}+\gamma'$$

1 mp alatt átvihető ennek reciproka: $v = \frac{C}{H} - \delta$

(Bizonyos késleltetés van, míg elő áll az adott hosszú sorozat)

— o —

Nevezetes kódolás:

(hibajavító - Hamming)
(titkosítási - RSA)

Huffman kód - tömörítési jelleggel

Előre meg kell mondani mit mivel kódolunk - a jövőben véletlen eloszlás szerint érkeznek a szimbólumok.

$X_1, \dots, X_n$ - lehetséges szimbólumok

Bináris kód: $X_i \to w_i$ bin kód

$l(w_i)$ - kód hossza

Feladat: visszafejthetőség

$X_{i_1}, \dots, X_{i_N} \to \underbrace{w_{i_1} \dots w_{i_N}}$

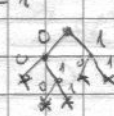
ebből az eredeti sorozat visszaállítható legyen

Prefix kódolást használva garantált a visszafejthetőség (Prefixmentes változatban)

Azaz: $\forall i,j$-re w_i nem hosszabbik része w_j-nek

$\sum_{i=1}^{n} 2^{-l(w_i)} = 1$ (véges teljes kódrendszer)

Kódfa:



Általánosan: véges teljes prefix-mentes Kraft-egyenlőtlenség:

$$\sum_{i=1}^{\infty} 2^{-l(w_i)} \le 1$$

A szimbólumok eloszlása: $(P_1, \dots, P_n)$

$X_1 - P_1$ üzenet választom

$\vdots$

$X_n - P_n$

Kódhosszak várható értéke: $\sum_{i=1}^{n} P_i \cdot l(w_i)$

Tétel: $H(P_1, \dots, P_n) \leq \sum_{i=1}^{n} P_i \cdot l(w_i)$ minden prefix kódra

($B_i \geq 1$ indirekt + de akkor nagyobb sebességet érhetnénk el mint az alaptétel mond)

$B_i \geq 2$:

a_i, b_i, a, b

$$\sum_{i=1}^{n} a_i \log \frac{b_i}{a_i} \leq a \log \frac{b}{a}$$

$a_i \rightarrow (P_1, \dots, P_n) \quad (q_1, \dots, q_n) \leftarrow b_i$

$$\sum_{i=1}^{n} P_i \log_2 \frac{q_i}{P_i} \leq 1 \cdot \log 1 = 1 \cdot 0 = 0$$

$$H(P_1, \dots, P_n) \leq -\sum_{i=1}^{n} P_i \cdot \log_2 q_i$$

$$\sum_{i=1}^{n} 2^{-l(w_i)} = 1 \quad , \quad q_1 := 2^{-l(w_i)}$$

$$H(P_1, \dots, P_n) \leq -\sum_{i=1}^{n} P_i \log_2 2^{-l(w_i)}$$

$$H(P_1, \dots, P_n) \leq -\sum_{i=1}^{n} P_i \cdot (-l(w_i))$$

$$H(P_1, \dots, P_n) \leq \sum_{i=1}^{n} P_i \cdot l(w_i)$$

Nevezetes kódolások:

1) Schannon-Fano kód:

$$\sum_{i=1}^{n} P_i \cdot l(w_i) < H + 1$$

(1 bit veszteséggel tudjuk megközelíteni az entrópiát)

Kód elkészítése: $P_1 \geq P_2 \dots \geq P_n$

$Q_1 := 0$

$Q_i := \sum_{j=1}^{i-1} P_j \quad (i = 2, \dots n)$

Kód: binárisan kódoljuk a Q_i számokat,

$0, \alpha_1 \dots \alpha_{l_i}$ l_i pontossággal megadjuk

Úgy választjuk l_i-t, hogy: $2^{-l_i} \le P_i \le 2^{-l_i+1}$

($l_i - 1$ 0-ás bittel kezdődik)

$l_i \ge -\log_2 P_i \ge l_i - 1$ $\quad / \cdot P_i$ és Σ

$$-\sum_{i=1}^{n} P_i \log_2 P_i \ge \sum_{i=1}^{n} P_i (l_i - 1)$$

$$H \ge \left(\sum_{i=1}^{n} P_i l_i\right) - 1$$

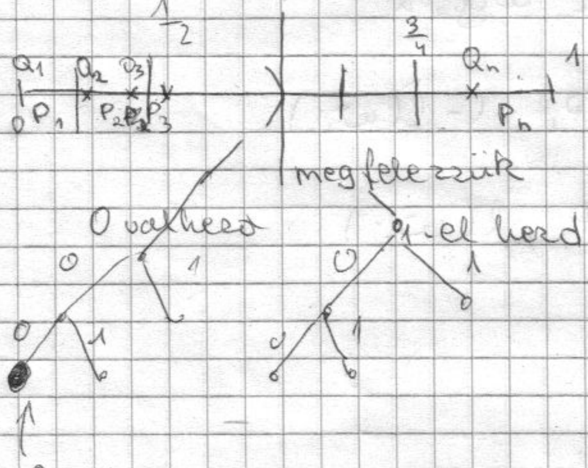
$$\sum_{i=1}^{n} P_i l_i < H + 1$$

prefix kód?

- hosszak nőnek
- „érték" nő

w_i w_{i+1} között teljesül a pref. mentesség?

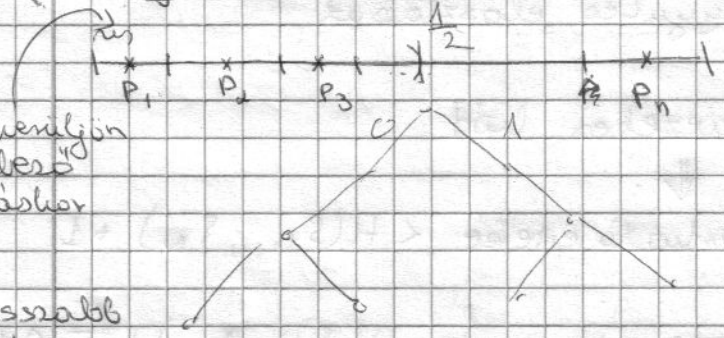
Igen, ui hozzáadunk az utolsóhoz valamit és így változik, nem lesz prefix.



leáll: az intervallumban egyetlen pont marad

(Hátránya: rendezni kell az eloszlásokat)

2) Gilbert - Moore kód : $p_1, \dots, p_n$ nincs rendezve



intervallum felező-pontjára: felezéses kódfa

ide kerüljön a felező pont beállításkor

⇓

1-el hosszabb a kód

leáll - ha egyenlő maradt

(1-el többször kell felezni: $\frac{p_i}{2}$ legyen benne)

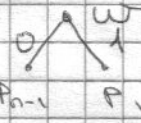
$$\sum_{i=1}^{n} p_i l_i < H + 2$$

3) Huffman - kód

Rekurzió: $p_1, \dots, \underbrace{p_{n-1}, p_n}$

a két legkisebb : p_{n-1}, p_n

kódja $w0$ p_{n-1}
$w1$ p_n



$\rightarrow q = p_{n-1} + p_n$

Aztán vesszük a $(p_1, \dots, p_{n-2}, q)$ Huffman kódját

w

Optimális: Bármely más prefix kódra w'_i:

$$\sum_{i=1}^{n} p_i l(w_i) \geq \underbrace{\sum_{i=1}^{n} p_i l(w_i)}_{\text{Huffman-kód}}$$

$$\Downarrow$$

$$\sum_{i=1}^{n} p_i l(w_i) < H + 1$$

Ha gyakoriság van: x_i – n_i - szer fordul elő

$$\sum_{i=1}^{n} n_i = N$$

$$\min \sum_{i=1}^{n} n_i \cdot l(w_i)$$

$$\Leftrightarrow \min \sum_{i=1}^{n} \frac{n_i}{N} \cdot l(w_i)$$

↑ eloszlás

Közelítsük $\frac{c}{H}$ hányadost:

Vegyük a $(\xi_1, \ldots, \xi_N)$ együttes eloszlását

$H(\xi_1, \ldots, \xi_N) \rightarrow$ Nevezetes kód

$\Downarrow$

Kódhossz várható értéke $< H(\xi_1, \ldots, \xi_N) + 1$

Egy szimbólumra jutó kódhossz: $\frac{1}{N} \cdot H(\xi_1, \ldots, \xi_N) + \frac{1}{N} \sim H + \frac{1}{N}$

(Tetszőlegesen közel tudok kerülni a $\frac{c}{H}$ sebességgel, ha nagyon sok hosszú kódszósorozatot kódolok egyben)

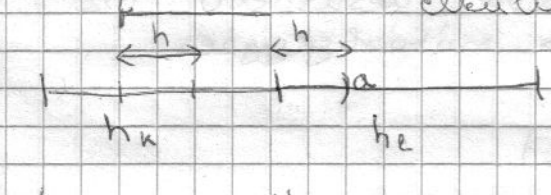
← "Nagyon nagy kódok, valamilyen tipikus halmazon való viselkedéshez hasonlóan működnek."

7. előadás, 2009.03.25., Infókc 4.

Tömörítés:

Lempel-Ziv kódok:

- LZ77 kód: (hivatkozok az előzőekben már elküldött kódokra)



már volt ezt akarjuk most küldeni

- minden szimbólumhoz kell (lesz) kód, azaz ha n szimbólum van – [log n] hosszal tudjuk kódolni

Átküldött kód: (t, h, c) , ahol c-az "a" kódja

$\ell(t) \sim \log_2 h_k$

$\ell(h) \sim \log_2 h_\ell$

↓ időtől függetlenné tehető" / időbeni és térbeli átlagolhatóság felcserélhető

Stacionárius, ergodikus forrás esetén – ezzel elérjük az optimális kódolást

Szótár kódok – szótárépítő kódolás

LZ78 kódolás:

- kezdetben ∀ szimbólumnak van egy kódja
- az új szövegrészben megnézi, milyen az előzővel leghosszabb egyezés van amivel egyezik, az már szótárban van és legyen a szótárkódja i
- Átküldi: <i, c> – ahol c a folytató "a" kódja
- ebből készül új szótárbejegyzés

Stac + ergo. forrás esetén – ennél is elérhető az optimalitás, azaz a max sebesség.

(de egy idő után a kiszélesítés nagy helyet)

LZW-kód (T. Welch)

c legyen "a" kódja

az előző <i, c> párból csak i-t küldjük, a következő kódolandó szó: "a"-val kezdődik

Majd a következő üzenetből készíthető az LZ 78 kód szerinti <i, c> szótárbejegyzés.

(Pl GIF is ilyet használ)

Korlátozzuk a szótárméretet!

— o — o — o — o — o — o — Kiszámítható

Kolmogorov-bonyolultság (-II- -entrópia) (1964) | Algoritmikus információelmélet (1960-as évek kezd

- Mint tömör leírása

(Sh: a kódkosszt minimalizálni)

(Kol: egy nagyon hosszú kódot, hogyan tudok tömörebbet előállítani, és hogy állítható ebből vissza)

- Egy nagy adat lehető legrövidebb tömörítése, amiből visszaállítható.

(Hogyan lehet a könyvet és Bélát tömöríteni, hogy egy számítógép ebből visszaadja a szöveget nyomtatva?
O-ból visszaállítja
Nem elég a kód, kell a dekódolás is)

Tömöríthetőség:

x tömöríteni, kód P, $\ell(P) = x$

X alakja	$\ell(P)$	x kódja	"hossz"
$x \in \mathbb{N}$	$f_0(P) = P$	x	$\ell(x) = \log_2 x$
$x = 2^k$	$f_1(P) = 2^P$	k	$\log_2 \log_2 x$
$x = 2^{2^k}$	$f_2(P) = 2^{2^P}$	k	$\log_2 \log_2 \log_2 x$
$x = 2^{2^{\cdot^{\cdot^{2}}}}$ (k db 2)	$f_3(P) = 2^{2^{\cdot^{\cdot^{2}}}}$ (P db 2)	k	$\underbrace{\log_2 \dots \log_2}_{k} x$

Ezeket a tömörítő fv-elvet használjuk.

x - tömörítése : i - melyik fv-t használom
p - kód (tömörítetten a legrövidebb)

(i, p)

$X = 1025 = 1024 + 1 = 2^{10} + 1$

Lehetne: $f_4(p) = 2^p + 1$

f - kiszámítható, és ezek megszámlálhatóan sokan vannak, tehát az f-ek halmaza rekurzívan felsorolható

Ilyen fv-ek építkezése:
1. építkezés - Turing gép

2. Parciális rekurzív fv-ek:

$f_v : \mathbb{N}^k \to \mathbb{N}$ $\quad \mathbb{N} := \mathbb{N}_0$

Alapfv.-ek:
- zéró fv.: $Z : \mathbb{N} \to \mathbb{N}$, $Z(x) = 0$
- rákövetkező: $S : \mathbb{N} \to \mathbb{N}$, $S(x) = x + 1$
- vetítő (projekció): $P_{n,i} : \mathbb{N}^n \to \mathbb{N}$, $P(x_1, .., x_i, .., x_n) = x_i$

végtelen sok fv.

Függvény képzések:

- kompozíció: $f : \mathbb{N}^n \to \mathbb{N}$ képződik
  $h : \mathbb{N}^k \to \mathbb{N}$
  $g_1, .., g_k : \mathbb{N}^n \to \mathbb{N}$

$$h(g_1(x_1, .., x_n), .., g_k(x_1, .., x_n)) =: f(x_1, .., x_n)$$

- Primitív rekurzió: $f : \mathbb{N}^{n+1} \to \mathbb{N}$
  $g : \mathbb{N}^k \to \mathbb{N}$
  $h : \mathbb{N}^{n+2} \to \mathbb{N}$

$$f(x_1, .., x_n, 0) := g(x_1, .., x_n)$$
$$f(x_1, .., x_n, y+1) = h(x_1, .., x_n, y, f(x_1, .., x_n, y))$$

Primitív rekurzív fv-ek osztálya a:
- zárt a kompozícióra, primitív rekurzióra ($Z, S, P_{n,i}$)
- generáló sorozat

Lehet adni olyan fv.-t ami nem primitív rekurzív, ezért Büvészmutyi meg!)

- A generáló sorozathoz könnyű grammatikát szerkeszteni (véges ábécével)
- sorszámozhatók a generáló sorozatok
- univerzális fv:

  $\underline{U(k,x)}$: - $\forall$ generáló sorozathoz van k, amire $u(k,x) = f(x)$

  - az $u(k,x)$ minden k-ra primitív rekurzív
- $u(k,x)$ kiszámítható
- $\boxed{u(k,k)+1 = h(x)}$ fv-t tekintsük
  -- kiszámítható
  -- $h(x)$ milyen fv?
  -- $h(x)$ nem primitív rekurzív
  -- ha az lenne, n lenne a sorszáma

  $h(n) \begin{cases} = u(n,n)+1 \\ = u(n,n) \end{cases}$ (hiszen n a h sorszáma)

  -- ellentmondás $\rightarrow$ $h(x)$ nem primitív rekurzív
- Pr. rek. fv. hátránya: lassan növekszenek
- Ackermann fv-ek

  $A : \mathbb{N}^2 \rightarrow \mathbb{N}$,

  $A(0,y) = y+1$

  $A(x+1, 0) = A(x,1)$

  $A(x+1, y+1) = A(x, A(x+1, y))$

  $a_m(x) = A(m,x)$

  a_4 : toronyhatvány, aminek a magassága x.

  a_5 : -"- , -"- $a_4^x(x)$

- Egy pr. rek. fv az $\forall$ valamelyik Ackermann fv-nél lassabban növekszik.
- $a(m) := a_m(m)$ - $\forall$ pr. rek. fv-nél gyorsabban növekszik

- Még egy képzés szükséges:

- minimalizációs fv-képzés:

$f: \mathbb{N}^n \to \mathbb{N}$ képződik

$g: \mathbb{N}^{n+1} \to \mathbb{N}$

$f(\vec{x}) = My([g(\vec{x}, y) = 0])$

az $f(x_1, .., x_n) = k$, ha

1) $g(x_1, .., x_n, k) = 0$

2) $y < k$: $g(x_1, .., x_n, y) \neq 0$

3) $g(x_1, .., x_n, y)$ értelmezve van $\forall y \leq k$

(Tehát kiértékelhető 0-tól k-ig is)

// Pl: $g(x,y) = x+y+1$

$f(x)$ - sehol nincs értelmezve

(Parciális fv-t kapunk)

- Minimalizációs használatával a parciális rekurzív fv.-eket kapjuk.

- És az osztály azonos a Turing-kiszámítható ~~fv-ek~~ fv-ek osztályával
(Church-tézis)

Szigorú változat: reguláris minimalizáció

- $\forall x_1, .., x_n$-re van $h: g(x_1, .., x_n, h) = 0$
- g totális fv.

$\Rightarrow$ Rekurzív fv-ek osztályát kapjuk.

Az univerzális fv. nem vezet ki a parciális rekurzív fv-ek osztályából. $g(x)$

$U(h, x) = g(x)$, ha g sorozima
↑ kapott, akkor értelmezve vannak

$h(x) = U(x, x) + 1$

Sorszáma : n

$U(n, x) = h(x)$

$h(n) \begin{cases} = U(n, n) + 1 \\ = U(n, n) \end{cases}$

ha $h(n)$ értelmezve lenne $\Rightarrow$ ellentmondás

$\Rightarrow$ $h(n)$ nincs értelmezve

(átlós fv-ek elsorolásra nem értékelhetőek)

- Ugyanez nem működik a regulárisra
  $\Rightarrow$ nincs univerzális felsoroló fv.-ünk

(regularitás eldönthetetlen algoritmikusan)
$\Rightarrow$ tehát nem lehet megkonstruálni a generáló sorozatokat

// Kiértékelhetőség esetén működik, hogy egy minimalizáló ciklus nincs befejeződik olyan helyen, ahol a fv nincs értelmezve.

2009.04.01, 8. előadás, Infókéz 4.

(XML-ről Ullman-Widow könyv végén)

Kleene-féle normál alak: $\exists U, V$:

$U: \mathbb{N} \to \mathbb{N}$, $V: \mathbb{N}^3 \to \mathbb{N}$ prim. rekurzív fv-ek

úgy, hogy bármely $f: \mathbb{N} \to \mathbb{N}$ parciális rekurzív függvényhez létezik n,

$$f(x) = U(\mu t\,(V(n,x,t) = 0)) \quad \forall x\text{-re}$$

$\overset{\circ}{\mathbb{R}}$

$A \subseteq R$

$x \in A$ véges időben megmondható

$$h_A(x) := \begin{cases} 0, & \text{ha } x \in A \\ 1, & \text{ha } x \notin A \end{cases}$$ (ha h totális fv)

A-rekurzív (eldönthető), ha h_A rekurzív

$A \subseteq R$ rekurzívan felsorolható, ha létezik $f: \mathbb{N} \to A$ rekurzív fv., hogy $x \in A \Leftrightarrow \exists n : f(n) = x$

Ha $x \in A$ - véges lépésben bizonyítható,
ha $x \notin A$ - tudjuk-e igazolni

A rekurzív $\Leftrightarrow$ A és $\neg A$ is rekurzívan felsorolható

Rice-tétel: (U - univerzális fv. szerinti sorszámozás)

"$\mathcal{A}$" tetszőleges részhalmaza a parc. rekurz. fv-eknek

$A = \{n \mid f_n \in \mathcal{A}\}$, ekkor A a.cs.a. eldönthető

ha $\mathcal{A} = \emptyset$ v. $\mathcal{A}$ = teljes

(Ilyen tulajdonság pl., h. a Turing-gépes probléma megoldható, vagy hogy egy parc. rek. fv rekurzív-e)

Univ. Turing gép - i-edik Turing gépet adok és szimulálja, ha i-t adok neki paraméternek

Kolmogorov - bonyolultság: (Algoritmikus entrópia)

(Kolmogorov, Chaitin, Solomonoff) ~ 1964

Leképezést készítsünk:

$\Omega = \{0,1\}$, $\mathbb{N} = \{0,1,\dots\}$

Λ - üres szó	0
0 = } 1 bit hosszú	1
1 = }	2
00	3
01	4
10	5
11	6
000	7
001 - - - - - - - -	8
010	
011	
100	
101	
110	
111	

(100 - lexikografikus sorrend)

Kölcsönösen egyértelmű

$x \in \Omega$ / $\in \mathbb{N}$ $\quad l(x)$ - x hossza ($\approx \log_2 x$)

Konkatenáció: xy

Rendezett párok kódolása: $(x,y) \to z$

①

	0	1	2	3	4
0	0	2	5	9	
1	1	4	8		
2	3	x			
3	6				

(x,y) hossza nagy lehet, bár x pl. kicsi

② Önkorlátozó kódolás: $\bar{x}$ (prefix lehet az $\mathbb{N}$ kódból)

$x = a_1 a_2 \dots a_n$

$\bar{x} = \underbrace{11\dots1}_{n} 0 a_1 \dots a_n$

$l(\bar{x}) = 2l(x) + 1$

$x' = \bar{n}\, x \qquad l(x') = 2\log_2 n + l(x) + 1$

$2\log_2(l(x)) + l(x) + 1$

(Lehetne folytatni $\bar{n}$ helyett n', stb...)

(x,y)

$\tau_0(x,y) = \overline{xy} = z_0 \xrightarrow{\text{visszafejtés}} \overline{\Pi}_{0,1}(z_0) = x_1,\ \overline{\Pi}_{0,2}(z_1) = y$

$\tau_1(x,y) = x'y = z_1 \qquad \overline{\Pi}_{1,1}(z_1) = x,\ \overline{\Pi}_{1,2}(z_1) = y$

— ∘ —

X - mennyire tömöríthető?

$f: \mathbb{N} \to \mathbb{N}$

Def: Az $x \in \mathbb{N}$/ra bonyolultsága az $f: \mathbb{N} \to \mathbb{N}$ függvény szerint

$C_f(x) = \min\{\ell(p) \mid f(p) = x\}$, ∞, ha nincs p.

(C-bonyolultság)

Alaptétel: (Invarianciatétel) (K-Ch-S)

Létezik olyan optimális $f_0: \mathbb{N} \to \mathbb{N}$ par. rek., hogy $\forall f: \mathbb{N} \to \mathbb{N}$-hez létezik k_f konstans, hogy $C_{f_0}(x) \le C_f(x) + k_f \quad \forall x \in \mathbb{N}$-re.

Következménye:

f_0, g_0 is optimális

$|C_{f_0}(x) - C_{g_0}(x)| \le k_{f_0,g_0}$

Optimális fv-ek szerinti bonyolultság konstans erejéig egyértelmű

$\Rightarrow$ Rögzítsük az f_0 optimális fv-t.:

Def: Az $x \in \mathbb{N}$ Kolmogorov bonyolultsága (entrópiája)

$C(x) = C_{f_0}(x)$

(tömöríthetőség elvi mérőszáma)

Alaptétel bizonyítása:

Legyen x, $f: \mathbb{N} \to \mathbb{N}$ tetszőleges

$C_f(x) = k$, létezik $f(p) = x : \ell(p) = k$

Univerzális $U(y,n)$ fv-t, ahol n: -sorszám

f-hez $\exists n: U(y,n) = f(y)$

z-t választom úgy, hogy

$z := n^\frown p$ - ből

$f_0(z) = U(\Pi_{1,2}(z), \Pi_{1,1}(z)) = U(p,n) = f(p) = x$

$\Rightarrow C_{f_0}(x) \leq l(z) = l(n') + l(p) = l(p) + k_f = C_f(x) + k_f$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \overset{\shortparallel}{k_f}$

// Eljárás leírásának hossza : $l(n')$
kódleírás -"- : $l(p)$ } $\Rightarrow$

// Tömörítésnél $\Rightarrow$ minimalizáljuk a leírás és a visszaírás hosszát)

Alaptulajdonságok $C(x)$ -re:

1. Mit jelent az, hogy $C(x) = k$?
$\Rightarrow$ Létezik egy p kód ; $f_0(p) = x$ és $l(p) = k$

1.a. $C(x) \leq k$? $f_0(p) = x$ és $l(p) \leq k$
(2^{k+1} ilyen kód lehet)

2. Alaptétel: felülről becsülhető adott $f: \mathbb{N} \to \mathbb{N}$
$C_{f_0}(x) \leq C_f(x) + k_f$

3. Van-e minden x-nek bonyolultsága?
$C(x)$ totális függvény // (2)-es felhasználásával
$f(p) = p$ -t használva
$C(x) \leq C_f(x) + k_f = l(x) + k_f$

(A Kolm. bony. max konstanssal nagyobb a hossznál)
"Semmisem bonyolultabb, mint amilyen"

4 $C(x)$ nem kiszámítható
$(C_f(x)$ - ha f nincs értelmezve a p helyen, $l(p) = k$
$\Rightarrow$ nem tudom megmondani, h. p semminek nem a kódja

Minimalizáció - nem tudom megmondani, hogy nem kisebb-e k-nál valami)

Indirekt: $C(x)$ kiszámítható:

x_0 : az első x, amire $C(x) \geq 1 = 2^0$
x_1 : az első x_0 után, amire $C(x) \geq 2$

4/8

:

x_k az első x_{k-1} után, amire $C(x) \geq 2^k$

Legyen:

$f(p) = xp$

$C_f(xp) \leq l(p) + k_f$ } Inv. tul:

$C(x_p) \geq 2^p$ $\quad C(x_p) \leq l(p) + k_f$

$2^p \leq l(p) + k_f$ | ellentmondás

$\Rightarrow$ nem kiszámítható

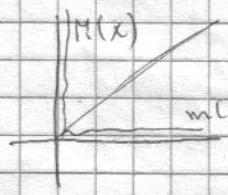
5. $\lim_{x \to \infty} C(x) = +\infty$ // (1)-es alapján

6. $m(x) := \min_{x \leq y} C(y)$ ($C(x)$ alsó burkolója)

$\lim_{x \to \infty} m(x) = +\infty$

$m(x)$: $\forall$ monoton kiszámítható fv-nél kisebb

(Indirekt bizonyítással)



$M = m(x)$ ~~is~~ tükrözöttje az $x=y$-ra

Ezek között van a bonyolultság

7. $(x+h)$-t előrészolni

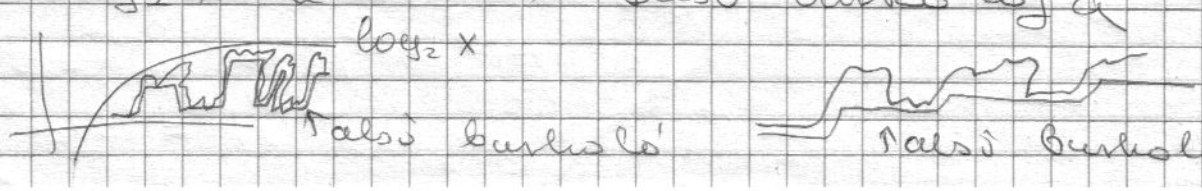
$C(x+h) \leq C(x) + 2 \cdot C(h) + k_+$

(kicsit változtatok az x-en kicsit változik a bonyolultság)

$g(x,h) = y$ fut használjuk

$C(g(x,h)) \leq C(x) + 2C(h) + \underbrace{k_g}_{\text{változtatás kiválki értéke}}$

8. $\log_2 x$ a $C(x)$ felső burkolójára



2009. 04. 22, Infókéz 4., 10. előadás

- Ullman könyv utolsó két fejezete XML

XSLT - Extensible Stylesheet Language for Transformations

// Dokumentumok átszerkesztését biztosítja

XSLT-stíluslap:

Stíluslapokat adunk meg, ezek is XML dokumentum formátumúak

xsl: jelölő - névtérhez kötődik

```
<?xml ... >
<xsl:stylesheet xmlns:xsl="http://www.w3.org/..."
  ⋮ (stylesheet kitöltése)
</xsl:stylesheet>
```

Stíluslap - sablonokból áll (template)

sablon jelölője: xsl:template

speciális attribútumai:

- match (milyen elemekre kell majd alkalmazni)

```
<xsl:template match="XPath kifejezés">
  ⋮ (sablonon belül, amit elő akarok állítani,
     ez kerülhet a template alkalmazásakor)
</xsl:template>
```

XPath itt lehet - abszolút: /
- relatív: egy elem feldolgozásán belül használjuk

Templaten belül egy speciális feldolgozás lehetőség, az elem gyerekeire kérni feldolgozást:

```
<xsl:apply-templates select="kifejezés"/>
```

- olyan templatet keres, aminek a match feltételére illeszkedik a gyerekre
- relatív útvonallal lehet megadni a matchet

(X) 10.

Ha illeszkedik a select feltételkifejezésre, akkor hajtja végre a templatet.

- Rekurzió hívást is lehetővé tesz...

XML-ből értékek elérése:

<xsl:value-of select="kifejezés" />

kifejezés - XPath, eredménye szöveg-nek tekintendő

Iteráció:

<xsl:for-each select="kifejezés">
⋮ (- amit végre kell hajtani)
</xsl:for-each>

Feltételes parancsok:

<xsl:if test="logikai kif."> ← then
[törzs - mit akarunk ennek hívni]
</xsl:if>
<xsl:if test="negált log. kif"> ← else
⋮
</xsl:if>

// Van choose is (otherwise ággal)
output (html, szöveg, XML)
↑ez alá a node alá kerül az eredmény

- Bejárás az XML adatain és a legjobban illeszkedő templatet választja.

```
1)  <? xml version = "1.0" encoding = "utf-8" ?>
2)  <xsl:stylesheet xmlns:xsl =
3)          "http://www.w3.org/1999/XSL/Transform">

4)      <xsl:template match = "/Filmek">
5)          <Filmek>
6)          <xsl:apply-templates />
7)          </Filmek>
8)      </xsl:template>

9)      <xsl:template match = "Film">
10)         <Film filmcím = "
11)         <xsl:value-of select = "@filmcím" />
12)         ">
13)         <xsl:apply-templates />
14)         </Film>
15)     </xsl:template>

16)     <xsl:template match = "Verzió">
17)         <xsl:apply-templates />
18)     </xsl:template>

19)     <xsl:template match = "Színész">
20)         <Színész név = "
21)         <xsl:value-of select = "." />
22)         " />
23)     </xsl:template>

24) </xsl:stylesheet>
```

12.24. ábra. A `filmek.xml` fájl transzformálása

12.22. példa. Ebben a példában az XSLT-t arra fogjuk használni, hogy XML-dokumentumot egy másik XML-dokumentumba transzformáljuk ahelyett, hogy egy HTML-dokumentumba transzformálnánk. Vizsgáljuk meg a 12.24. ábrát. Négy sablont találunk benne, és ezek együtt fogják feldolgozni a 12.22. ábra filmadatait. Az első sablon, a 4. és 8. sorok között a gyökérre illeszkedik. Ez azt mondja, hogy a kimenet legyen a `<Filmek>` szöveg, majd pedig alkalmazzuk a sablont a gyökérelem összes gyerekére. Specifikálhatnánk, hogy a sablont csak azokra a gyerekekre alkalmazzuk, amelyek a `<Film>`-mel vannak jelölve, és semmilyen más jelölőt nem várunk a gyerekek közül, ezért nem így specifikáltuk:

```
6)      <xsl:apply-templates select = "Film" />
```



```
1)  <? xml version = "1.0" encoding = "utf-8" ?>
2)  <xsl:stylesheet xmlns:xsl =
3)          "http://www.w3.org/1999/XSL/Transform">
4)      <xsl:template match = "/">
5)          <HTML>
6)              <BODY>
7)                  <B>Ez egy dokumentum</b>
8)              </body>
9)          </html>
10)     </xsl:template>
11) </xsl:stylesheet>
```

12.21. ábra. Nyomtatási kimenet tetszőleges dokumentumokhoz

```
<? xml version="1.0" encoding="utf-8" standalone="yes" ?>
<Filmek>
    <Film filmcím = "King Kong">
        <Verzió év = "1933">
            <Színész>Fay Wray</Színész>
        </Verzió>
        <Verzió év = "1976">
            <Színész>Jeff Bridges</Színész>
            <Színész>Jessica Lange</Színész>
        </Verzió>
        <Verzió év = "2005" />
    </Film>
    <Film filmcím = "Gumiláb">
        <Verzió év = "1984">
            <Színész>Kevin Bacon</Színész>
            <Színész>John Lithgow</Színész>
            <Színész>Sarah Jessica Parker</Színész>
        </Verzió>
    </Film>
        ... további filmek
</Filmek>
```

12.22. ábra. A `filmek.xml` fájl

12.3.3. XML-adatokból elérhető értékek

Szokatlan, hogy a kimenetként kapott dokumentum semmilyen módon nem függ a transzformáció bemenetétől, ahogy az a 12.20. példa esetében történt. A leg-

Figyeljük meg, hogy `<Film>` gyerekekre vonatkozó sablono után (ami az eredményben több elem kiírását eredményezi), a kim tuk a `<Filmek>` elemet a megfelelő záró jelölővel a 7. sorban. Fig is, hogy meg tudjuk különböztetni a kimenetben szereplő jelölőke az 5. és 7. sorokban vannak –, az XSLT-jelölőktől, mivel min jelölőnek az `xsl` névtérből kell származnia.

Most nézzük meg, hogy a `<Film>` elemekre milyen sablonokat Az első (és egyetlen) sablon, amely ezekre az elemekre illeszked sablon, amely a 9. és 15. sorok között található. Ez a sablon `<Film filmcím= "` szöveggel kezdi a 10. sorban. Ez követően a l a film címét és átadja azt a kimenet számára. A 12. sor befejezi attribútumot és a `<Film>` jelölőt a kimenetben. A 13. sor alkalma a film összes gyerekére, amelyek a verziók lehetnek. Végezetül a `</Film>` záró jelölőt.

Amikor a 13. sor meghívja a sablont, hogy alkalmazza ezt verziójára, az egyetlen illeszkedő sablon a 16. és 18. sorok között kívül nem tesz semmit azért, hogy a sablont alkalmazza a verzió kére, amelyek a `<Színész>` elemek lehetnek. Így az, hogy mi ker az egyes `<Film>` jelölők és a nekik megfelelő záró jelölő között, a rok közötti utolsó sablon határozza meg. Ez a sablon minden egy elemre vonatkozik.

A bemenet színészelemeit transzformáltuk a kimenetre. Ahe színészek nevei szövegek lennének – ahogy a 12.22. ábrán –, a 19. dő sablon előállít egy `<Színész>` elemet a névvel, mint attribútu sor azt mondja, hogy válasszuk ki a `<Színész>` elemet magát (a „önmaga", ahogy az XPath-ban is), mint értéket a kimenet szám a teljes kimenet szöveg, így az elemek jelölői nem részei a kimenet mény pontosan az, amit akartunk, azaz a `név` attribútum értéke nem egy elem. Az üres `<Színész>` eleme a 22. sorban lesz teljes. bemenet a 12.22. ábrán adott, akkor a kimenetet a 12.25. ábra m

12.3.5. Iterációk XSLT-ben

A sablonokba ciklusokat is elhelyezhetünk, amelyek szabadságot nekünk abban, hogy egy elem meglátogatott gyerekelemei közül m mazzuk a sablont. A for-each ciklus az alábbi alakú:

```
<xsl:for-each select = "kifejezés">
```

A kifejezés egy XPath-kifejezés, amelynek értéke tételek szekvenci álljon a nyitó `<for-each>` jelölő és a neki megfelelő záró jelölő köz végrehajtásra kerül ciklusléptenként az egyes tételekre.


```
<xsl:value-of select = "kifejezés" />
```

A kifejezés egy olyan XPath-kifejezés, amely egy szöveget ál Egyéb értékek esetén – például amikor egy elem tartalmazza a vánvaló módon ezt szöveggé kell alakítanunk.

12.21. példa. A 12.22. ábrán reprodukáltuk azt a `filmek.x` a 12.2. alfejezetben használtunk példaként. Ebben a stíluslapr a `value-of` műveletet fogjuk használni arra, hogy elérjük az és kiírjuk azokat soronként egyesével. A stíluslapot a 12.23. á

A 4. sorban azt látjuk, hogy a sablon minden `Film` elem egyesével feldolgozhatjuk azokat. Az 5. sorban a `value-of` mű zuk egy `@filmcím` XPath-kifejezéssel. Ez azt jelenti, hogy egyes `Film` elem `filmcím` attribútumához és szerezzük meg az attribútumnak. Ez az érték kerül kimenetként előállításra, ma sorban egy HTML „új sor" jelölő található, így a következő fil sorban kerül kiírásra. □

```
1)  <? xml version = "1.0" encoding = "utf-8" ?
2)  <xsl:stylesheet xmlns:xsl =
3)          "http://www.w3.org/1999/XSL/Transfo
4)      <xsl:template match = "/Filmek/Film">
5)          <xsl:value-of select = "@filmcím" />
6)          <BR/>
7)      </xsl:template>
8)  </xsl:stylesheet>
```

12.23. ábra. A filmcímek kiírása

12.3.4. Sablonok rekurzív használata

A legérdekesebb és legerősebb transzformációkat várhatóan a zív alkalmazása jelenti a bemenet különböző elemeire. Egy kiv alkalmazva a bemeneti dokumentum gyökerére kérhetjük, ho egy sablont annak minden gyerekelemére is az `apply-templ` nálatával. Amennyiben egy kiválasztott sablont csak a gyerek halmazára – például azokra, amelyek egy kiválasztott jelölőv szeretnénk alkalmazni, használhatjuk a `select` kifejezést, mé

```
<xsl:apply-templates select = "kifejezés"
```

Amikor egy ilyen jelölővel találkozunk a sablonban, megker elem illeszkedő gyerekelemeit (az elemeket, amelyekre a sablo

(Vissza a bonyolultság elméletbez)

Példa:

$$\{ x \mid l(x) = n \wedge C(x) \leq n - \Delta \}$$

(legalább Δ-val jobban tömöríthetőek, mint amilyen hosszúak)

Mekkora lehet ennek az elemszáma?

A lehetséges kódok, amit fo használhat hossz szerint számoljuk le:

$$\begin{array}{r} 0 - 1 \\ + \quad 1 - 2 \\ + \quad 2 - 2^2 \\ + \quad \vdots \\ + \quad 2^{n-\Delta} \\ \hline 2^{n-\Delta+1} - 1 \end{array}$$

$C(x) \leq n - \Delta$, azt jelenti tömöríthető $n-\Delta$ hosszú kóddal

max elemszám

Hogyan aránylik 2^n-hez : $2^{-\Delta+1}$

„A dolgok többsége olyan bonyolult amilyen"

„ $\Delta = 2$: a hosszan max lehe tömöríthető $n-2$ hosszú kóddal"

Változik az x, hogyan változik a Bony.?

$y = x + h$

$C(x)$

$C(y)$ -?

ui. rendezett párt kell kódolni

$$g(x,h) = y$$

$$C(y) \leq C(x) + C(h) + 2 \cdot \log_2 C(h) + k_g$$

Biz: Invariancia tételre alapul, kell adnunk y-ra egy ügyes kódolást

$C(x) = k$, $C(h) = m$

$fo(p) = x$ $\quad$ $fo(q) = h$

$l(p) = k$ $\quad$ $l(q) = m$

p és q ismeretében ki kell számolni y-t! De nekünk nem a pár, hanem egy kód kell...

→ standard invertálható kód

$\tau = \langle q, p \rangle$

$$f(\tau) = g\Big(\underbrace{f_0(\underbrace{\Pi_{12}(\tau)}_{p})}_{x}, \underbrace{f_0(\underbrace{\Pi_{11}(\tau)}_{q})}_{h}\Big) = y$$

$$C_f(y) = l(\tau) = l(p) + l(q) + 2 \cdot l(l(q)) =$$
$$= k + m + 2 \cdot \log_2 m =$$
$$= C(x) + C(h) + 2 \log_2 C(h)$$

„Kicsit változik x, kicsit változik a bonyolultság"

Feltételes Kolmogorov-bonyolultság:
(relatív)

Mennyit segít y kiszámításához az x ismerete?
Ha ismerek egy $f: \mathbb{N}^2 \to \mathbb{N}$ fv-t: $f(x,p) = y$, akkor elég a p ismerete.

Def: Az $f: \mathbb{N}^2 \to \mathbb{N}$ felhasználásával az y feltételes bonyolultsága x ismeretében:

$$C(y|x) = \{ \min_{f(x,p)=y} l(p), \text{ illetve } \infty, \text{ ha nincs ilyen } p$$

Kérdés: van-e ebben az esetben is tömörítő f?

Alaptétel-invariancia tétel: létezik optimális $f_0^{(2)}: \mathbb{N}^2 \to \mathbb{N}$ parciális rekurzív fv, hogy bármely $f: \mathbb{N}^2 \to \mathbb{N}$ fv-hez megadható egy k_f konstans, hogy

$C_{f_0}(y|x) \leq C_f(y|x) + k_f \quad \forall x,y \in \mathbb{N}$-re.

(k_f nem függ x-től, csak f-től)

Biz: Kétváltozós fv-ek univerzális parc. rek. fv-t kell használni: $U(n, x, p)$

(ua. a konstrukció mint a leszámlálóeset esetben)

Invariancia-tétel: $f_0^{(2)}$, $g_0^{(2)}$ optimális:

$$\left| C_{f_0}(y|x) - C_{g_0}(y|x) \right| \leq k_{f_0, g_0}$$

- konstans erejéig megegyeznek.

Rögzítjük $f_0^{(2)} : \mathbb{N}^2 \to \mathbb{N}$ optimális fv-t.

Def: $C(y|x) = C_{f_0}(y|x)$

2009. 04. 29., 11. előadás, Infóhöz 4.

Feltételes Kolmogorov bonyolultság:

$$C_{f_0^{(2)}}(y|x) = C(y|x)$$

Alaptulajdonságai:

- invariancia-tétel alapján: $f: \mathbb{N}^2 \to \mathbb{N}$

$$C(y|x) \leq C_f(y|x) + k_f$$

↑ x, y-tól független konstans

- $C(y|X) = k$ jelentése: $\exists p: f_0(p, x) = y$ és $l(p) = k$
- $C(y) \leq C(y|x) + C(x) + 2 \cdot \log_2 C(x) + k$

(hasonlóan mint: $y = g(x,n)$ inv. tétellel $C(y)$-ra)

Mire használható? - Halmaz elemének bonyolultsága a

- Halmaz szerinti feltételes bonyolultság

$A = \{a_1, \ldots, a_N\}$

$C(a|A)$ - ? $\quad a \in A$

$\langle A \rangle = a_1 a_2' \ldots a_{N-1}' a_N^*$ ← A kódolása önkorlátos kóddal

$$C(a|A) = C(a|\langle A \rangle)$$

Adott A, $|A| = N$ Mennyire tömören ábrázolható?

$|\{x \mid x \in A, C(x|A) < \log_2 N - \Delta\}|$ - ? Mi az elemszám

Kány olb adott bonyolultságú elem lehet?

Lehetséges p kódok: $f_0(p, \langle A \rangle) = x \in A$

0 1 2 ... $\log_2 N - \Delta - 1$ ← kódhossz

$1 + 2 + 2^2 + \ldots + 2^{\log_2 N - \Delta - 1} = 2^{\log_2 N - \Delta} - 1$

$$\frac{|\{x \mid x \in A, C(x|A) < \log_2 N - \Delta\}|}{|\{x \mid x \in A\}|} = \frac{2^{\log_2 N - \Delta}}{2^{\log_2 N}} \leq 2^{-\Delta}$$

A halmaz is mesterében sem lehet tömörebben ábrázolni az elemeket az egyenletes kódolásnál.

az ↑ a legtöbb elemet

Keresem a lehető legkisebb halmazt, ami lefedi a jelenségem és egyenletesen kódolom.

Példa: Személyek adataira relációs AB:

R(csnév, unév)

csnév: 20 byte = 160 bit

knév: 20 byte = 160 bit

$\mathcal{A}$: az összes lehetséges relációelőfordulás

- lehetséges különböző sorok száma: $2^{160} \cdot 2^{160} = 2^{320}$
- lehetséges előfordulások száma: $2^{2^{320}}$ (részhalmaza)

$|\mathcal{A}| = 2^{2^{320}}$

$\{I \mid I \in \mathcal{A}\}$, $C(I \mid \mathcal{A}) = \log_2 2^{2^{320}} = 2^{320}$

Sorok számára van egy korlát a való életben:
< 10 milliárd

10 milliárd · 320 bit elég kódolni a sorokat

Példa: csnév: első 256 leggyakoribb
unév: -||-
Magyar statisztika szerint a populáció 80%-nak az a cs neve illetve 80%-nak az az utóneve.

4. fregmens:

1. Mindkét név gyakori : 64% - 2 byte
2. Csnév gyakori : 16% 22 byte
3. Unév gyakori : 16% 22 byte
4. egyik sem : 4% 4 byte

(Jól kell homogén részhalmazokra szétszedni)

Paraméteres halmazseregek:

(pl. adatbázis (relációk) sémával rögzítek egy paramétert és az első fordulás a halmazseregy)

$a \longrightarrow B_a$ véges halmaz
$h(a) = B_a$ – h egy totális fv (rekurzív) } megtudom mondani mi van benne és mi nincs

Legyen $A \subseteq N \times N$ rekurzívan felsorolható részhalmaza
$B_a = \{ x \mid (x, a) \in A \}$

megszorítás: minden a-ra $|B_a| = m_a$ véges
m_a-t nem tudom megmondani, mert nem tudom megmondani mi nincs benne

Tétel: Létezik k_A, hogy $\boxed{C(x|B_a) = C(x|a) \leq \log_2 m_a + k_A}$
$\forall x \in B_a$-ra és $\forall A$-ra

A halmaz elemeinek a tömöríthetősége mindig eléri a $\log_2$(elemszám) -ot.
(Alulról + felülről is becsülni tudjuk)

Biz:

B_a halmaz elemei:
A-hoz van felsoroló függvény:
$f(a, p)$: konstruáljuk
Futtatjuk a felsorolását, kapjuk az első $x_1 \in B_a$-t
$\Rightarrow f(a, 0) = x_1$
Tovább: $x_2 \in B_a$-t megkaptuk: $f(a, 1) = x_2$
⋮
$x_{m_a} \in B_a$ elem: $f(a, m_a) = x_{m_a}$
ez után az $f(a, p)$ nincs értelmezve

Kell: $C_f(x|a)$ – Kód egy $(0, \dots, m_a - 1)$ közötti érték.
hossza $\log_2 m_a$-val felülről becsülhető.

Érdekes halmazsereg:

$D_k = \{ x \mid C(x) \leq k \}$ k paraméterben ez egy halmazsereg...

Meg tudjuk-e konstruálni a felsoroló fv.-t?

fo optimális bo: hely és lépésszám bejárása...

fo: 0

fo: 1,

k \	0	1	2	3
1	1	1	2	3
1	1	2	3	
2	2	3		
3	3			

Ha nem ért véget veszem a leíró sztkót...

ha $x \in D_k$: ha valamelyik k hosszú értékre halunk $fo(p) = x$ eredményt.

—o—

$D_k^+ = \{ x \mid C(x) = k \}$ nem kiszámítható halmaz (ui. akkor a $C(x)$ is kiszámítható lenne, de általában nem lehet bebizonyítani, h. a halmog. bonyolultsága valamilyen k)

$D'_k = \{ x \mid C(x) < k \}$ ez sem kiszámítható

Nem tudom megmondani nincs-e jobb hívó nálam...

—o—

Kolmogorov-entrópia — Schannon-entrópia viszonya

Nagyon hosszú jelenségekre

Shannon-entrópia: N hosszú üzenetek halmaza (stacionárius, ergodikus forrás)

Két halmazba sorolható:

- lényeges, tipikus halmaz: ebbe $1-\varepsilon$ vsz.gel esik a választás

$$(1-\varepsilon) 2^{N(H-\delta)} < m_N < 2^{N(H+\delta)}$$

(m_N: elemszám)

$\bar{x}$ tipikus: statisztikus jellemzői B_N halmazba esik (ez lesz a par. halmazsereg)

$$C(\bar{x} \mid B_N) = C(\bar{x} \mid N) \leq \log_2 m_N + k_{\text{forrás}}$$

N hosszú üzenetek hosszát osszuk N-nel:

Shannon-entrópia: H

$$H \sim \frac{\frac{K_{olm}}{N \cdot (H-\delta)}}{N} < \frac{C(\bar{x} \mid B_N)}{N} < \frac{N \cdot (H+\delta)}{N} \sim H$$

Ha nagyon nagy jelenségeket vesszük, akkor a tip. halmaz Kolm. bonyolultsága visszaadja a Sch. entrópiát.

Hogyan: egyenletekkel jellemezhetőek

-o-

Összes közül melyek a tipikusak és melyek nem?

-o-

0-1 sorozatok véletlenszerűségét vizsgáljuk:

Lehet-e pénzfeldobás eredménye?

1000 pénzfeldobás: 00x10100 ... 1.. 101 -véletlenszerű

000x0100 ...

00000100 ...

egyeseket ~~nu~~ egyesével 0-ra cserélem:

:

00 ... 0 000 -nem véletlen

Van-e határvonal a véletlen és nem véletlen között?

Nem ...

Szabályszerűség

Ami igazán véletlen az nem tömöríthető ...

(Kezdőszelet bonyolultsága a hosszal nő!)

Mennyire tipikus?

Véletlenség definíciója jellemezhető: Δ

$$\{x \mid x \in A, C(x \mid A) < \log_2 |A| - \Delta\}$$

↑ véletlen definíciója a legalább Δ

Példa: Fekete - fehér színezések

(homogén ill. „mikus" sűrűség)

„sakktábla"

fele fekete – fele fehér

Egymásra helyezve a fóliát érdekes jelenség...

→ sajátgörbére illeszkedő transzformáció felelős...

2009. 05. 06., Infókez 4. 12. előadás

(Az információ nem növekedési törvénye!)

Adatbázis – x (x az adatbázis tartalma
Kérdés: y
Válasz: a = f(y, x)
Mennyi az a információtartalma?

$$C(a|x) \le C(y) + h$$

ui: $C(a|x) = C_f(a|x) + k_f \le C(y) + k_2$

Shannon-entrópia esetén:

$$-\sum_{i=1}^{n} p_i \log_2 p_i$$

p_i valószínűség – $-\log_2 p_i$ (mennyiséget rendeljük)
kódhossz l_i

x – C(x) (Kolmogorov esetben ez rendelj
(kódhossz)

$$p_i - 2^{-l_i}$$

Egész számokon adjunk meg egy v.sz. eloszlást:
$$P(i) = 2^{-C(i)}$$

$\sum_{i=0}^{\infty} P(i) = \infty$ (tehát nem jó eloszlásnak

Ha x-et rögzítem: $\sum_{fo(p)=x} 2^{-l(p)} = \infty$

összes lehetőség es kódra vesszük az összeget

Miért? Mert amit az fo felhasznál, nem alkot prefixmentes kódrendszert.

(1)/12.

f_0 -univerzális T-gép segítségével

$$U(i, p) = T_i(p)$$

Pénzfeldobással állítsuk elő a T-gép bemenetét, és legyen ez a p.
Nem jó, mert ...
Nem egyértelmű, hogy mire vezetett a pénzfeldobás

—o—

Prefix Kolmogorov-bonyolultság:
- Csak prefixmentes fv.-ekre használunk:

$f: N \to N$
$\Sigma^* \to \Sigma^*$
} ha az f értelmezve van p-re és q-n akkor egyik sem prefixe a másiknak

(Kraft-egyenlőtlenség: $\{p_1, \dots\}$ rendszer prefixmentes, akkor $\sum_{i=0}^{\infty} 2^{-l(p_i)} \leq 1$) (prefixmentes kódok)

Ha f prefixmentes, a $C_f(x)$ értelmezhető

Létezik optimális $g_0: \Sigma^* \to \Sigma^*$ prefixmentes fv., amire $\forall f$ prefixmentes fv.-hez létezik k_f konstans:

$$C_{g_0}(x) \leq C_f(x) + k_f$$

(ui. létezik univerzális felsoroló fv.)

Tetszőleges $h: \Sigma^* \to \Sigma^*$ parc. rek. fv.,
$\exists$ h_p: prefixmentes fv.
Idő, tér konstansnak tekintve futtatjuk a kiértékelést (párhuzamosan) és ha találtunk értelmezést akkor az tartozik a kódhoz, ha a következő, amit találunk prefixe az előzőnek, akkor kihagyjuk.

T_p - prefixmentes T-gép - véletlen sorozatok
k bit hosszú helyen, 2^{-k} valószínűséggel áll meg

$\sum 2^{-k} \leq 1$ a Kraft - egyenlőtlenség miatt.

$K(x) = C_{g_0}(x)$ prefixmentes Kolmogorov-bonyolultság!

$\pi(x) = 2^{-K(x)}$ (valószínűség)

$$\sum_{i=0}^{\infty} 2^{-K(i)} < 1$$

$$\Pi^*(x) := \sum_{g_0(p)=x} 2^{-l(p)} < 1$$

x - g_0 szerinti összes lehetséges kódja

g_0 - t képzeljük el T-gépnek, ahhoz annak a vsz-gc, hogy x-et ad éppen annyi

Azaz: a megállás valószínűsége x eredménnyel

(Azok fordulnak elő nagyobb vsz-gel, amit egyszerűbben tudunk ~~értelmezni~~ jellemezni)

$\pi(x)$, $\pi^*(x)$ - az egész számok felett értelmezett eloszlások

— o —

A legnagyobb bizonytalanság: egyenletes eloszlás.
Tudunk-e az egész számok halmazán egyenletes eloszlást megadni? Nem!
Mi a lehető leg-egyenletesebb eloszlás?
(Ami a lehető leglassabban csökken)

(3) Félig kiszámítható (semi) eloszlások között
12. a lehető legegyenletesebb eloszlás a $\pi(x)$ ($\pi^*(x)$)

Félig lisz.: nem tudom megmondani, de tetszőlegesen jól közelítem a pontos értéket.

Ha $p(x)$ félig lisz. eloszlás, akkor $\exists\, c$ konstans, hogy $p(x) < c \cdot \Pi(x)$
(Nem tud a $\Pi(x)$-nél gyorsabb lenni)

- Nagyorálú eloszlásnak nevezik ezt a tulajdonságot

(Algoritmikus tanulás hátterében ez van)

Probléma: nem tudjuk kiszámolni a $\Pi(x)$-et, csak közelítő értékekkel dolgozhatunk.

$C(x) = k \Rightarrow \exists p : f_0(p) = x$ és $l(p) = k$
$K(x)$
$g(p') = f_0(p) = x$
$l(p') = 2 \cdot \log_2 k + k$
$\Rightarrow K(x) \leq C(k) + 2 \cdot \log_2 C(k) + c$

—o—

Feltételes prefix-bonyolultság:
$f(x, p)$ p-ben prefixmentes

$C_f(y|x)$
$g_0^{(z)} : N^2 \to N$ optimális

$C_{g_0}(y|x) \leq C_f(y|x) + k_f$

$\Rightarrow$ $\underline{K(y|x)}$

Van értelme: $K(x|A)$, $x \in A$-nak

—o—

Shannon-entrópia esetében: $H(S, X) = H(S|X) + H(X)$
ez Kolmogorov-bonyolultságra:
(x, y) rendezett párra: $(x, y) = x'y$

$C(\langle x,y\rangle) = C(x'y)$

$C(x,y) \overset{?}{\leq} \underbrace{C(y|x) + C(x)}_{\text{túl nagy lehet}}$

És az additív tulajdonság nem teljesül, de K-ra

$$K(x,y) \leq K(x) + K(y|x) + 2\log_2 K(x)$$

Megj: Ha tudnám az x bonyolultságát, akkor konstans erejéig: $K(x,y) \overset{+}{=} K(x) + K(y \mid x, K(x))$

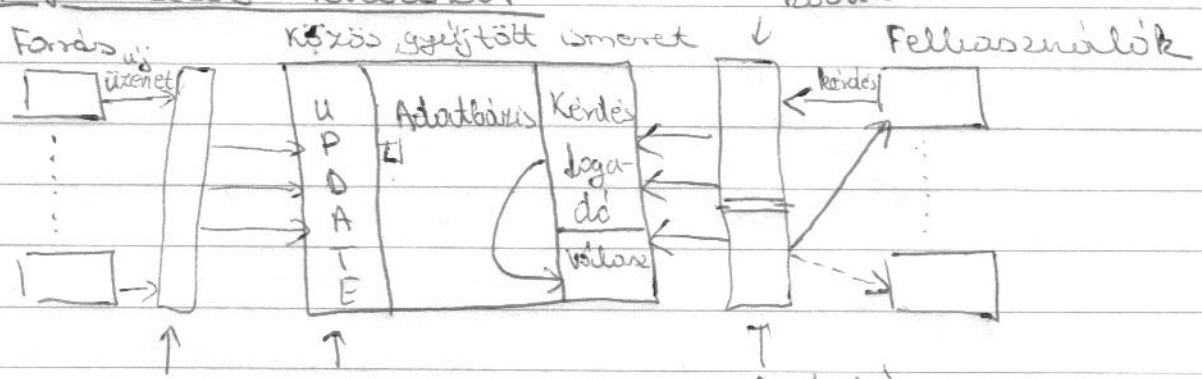
↑ additív

van x-nek tanúja:
x^* : $g_0(x^*) = x$
$l(x^*) = K(x)$

Elvi modell az adatbázis kereső rendszerekhez

Diagram: (Sch. modellhez hasonlóan)

Információs-rendszer



Kódolás Módosító felület dekódoló

Megszorítások:
- mi lehet az adatbázis tartalma? (valamilyen módon formalizálni kell)
- milyen új üzenetek lehetségesek, és hogyan épülnek be a régi ismeretekbe?
- milyen kérdések fogalmazhatók meg? válaszok?

 (tömöríthetőség, hatékony lekérdezés – ellentmond a módosíthatóság-nak)

13. előadás, 2009. 05. 13., Infókéz 4.

ABKR formális modellje:

1. logikai - fogalmi modell
Valóság modellezése - kódrendszer
↘ ↙
közötti kapcsolat

Célja:

Kódrendszer: kiszámíthatósági, bonyolultságelméleti elemzéssel

Valóság mod.: Absztrakt adattípusok adatmodellek előfordulásait → kódrendszer

Kódrendszer: bináris szavak rendszere

① Mi lehet az adatbázis lehetséges állapota a kódr. alapján?
↓ véges bináris szavak

$\Sigma \subseteq \Omega$ Σ - rekurzív v. rekurzívan felsorolható

Sémák: leírása adat, tehát:

$S \subseteq \Omega$

$s \in S$, Σ_s : séma szerint lehetséges adatok

$\langle s, y \rangle = x = s'y$ (s séma beli y előfordu

$s_1 \dots s_k$ sémák

$y_1 \dots y_k$ állapotok

$\langle s_i, y_i \rangle$ felsorolló kód

② Hogyan lehet módosítani az AB tartalmát?

Y - módosító adatok halmaza

a) $M : Y \times \Sigma \to \Sigma$ - módosító fv. kiszámítható fv.

b) M : módosító fv.ek halmaza

$m \in M$, y - lehetséges módosító adat ($y \in Y_m$)

$m(y, x) = x'$

- Zártsági elvárás : $M(Y \times \Sigma') \subseteq \Sigma'$ (szemantika adatmodellben dől el) generikus

olyan legyen a módosítás, h. az eredmény értelmes legyen

- ekvivalencia-osztályok bevezetése Σ-n

-üres módosítás : Λ, $x \equiv x'$ ha $M(\Lambda, x) = M(\Lambda, x')$

Sémánkénti módosítások: $s \in S$, Σ_s, $M_s \subseteq M$ - módosító fv. halm

Adat-mennyiségek:

- hossz (ezt mérhetjük) - redundancia
- $K(x)$, $C(x)$

→ ezek alapján elemezhető a módosítások komplexitása

Séma megválasztása mikor jó? (Kol. Bony. alapján)

$x \in \Sigma_s$, (Σ_s véges) $|\Sigma_s| = N$

$K(x|s) \sim \log_2 N$

Σ_s-en egyenletes kódot használhatunk

Σ_s-en homogén feladat lesz

Relációs séma: $S = R(A_1, \ldots, A_n)$

A_i – $|D_i| = m_i$

I – előfordulásra

lehetséges előfordulások száma: $2^{\prod_{i=1}^{n} m_i}$

sorok száma: $\prod_{i=1}^{n} m_i = m$ $\uparrow |\Sigma_S|$

Azokban: Tipikus feladat

Legfeljebb k sor van benne

$N - \binom{m}{k} \sim m^k$ – egyenletes kód: $k \cdot \log_2 m$

soronkénti kód ↑

(Soronként kódolják a táblákat a rel. AB-sok)

③ Lekérdezések: Q – lehetséges kérdések száma

$Q - q \in Q - \Sigma_q$ válaszhalmaz

típusos v. jól tipizált kérdésekkel foglalkozunk

∃ válasz fv q-hoz:

$$A: Q \times \Sigma \to \bigcup_{q \in Q} \Sigma_q$$

↑ kérdés ↑ állapot az AB-ből

$$A(q, x) = v \in \Sigma_q$$

- lehet, hogy a kérdésnek nincs értelme, de ezt tudom detektálni.
- lehet, h. nincs értelme, de ezt nem tudom megmon dani
- generikusság: ekv. állapotból ekv. választ kell kap

Fizikai modell

- bedarabolódik egy bináris kód címezhető, vegyes kapacitású rekeszekbe

↓hossz ↓adat

Ha van egy fizikai kód: (c_1, ℓ_1, y_1)

(c_2, ℓ_2, y_2)

$\vdots$

(c_R, ℓ_R, y_R)

φ - fizikai kód

φ , $z \in \Phi$

$\varphi: \Sigma \to \Phi$, $x \in \Sigma$, $\varphi_x \subseteq \Phi$

Kódosítások rendszere: $y \in Y$, φ_y

$w \in \varphi_y$, $z \in \varphi_x$

↓összefűzés

$\varphi_{x,y} = \varphi_y \circ \varphi_x = w z$

ℓ_M — díszjunkt címtartományon legyenek

$v \in \varphi_{m(y,x)}$

szgép megvalósítja implementálja a változást

Eljárásainknak is lesz egy kódolt formája, ugyanabban a címtartományban mozog.

Lekérdezés: q - $k \in \varphi_q$

$\varphi_q \; \varphi_x \; \varphi_{ABKR} \to \varphi_{A(q,x)}$ ($\varphi_x \; \varphi_{ABKR}$)

válasz

vagy

tárolódik (pl. materializált nézet)

dekódolás

$\Sigma_q \to$ felhasználói üzenet