

SZEMANTIKA, 2. ELŐADÁS

Kornai András

BME 2020/9/16

JELENTÉS ÉS INFORMÁCIÓ

- Mi az információ?
- Shannon modell (zajos csatorna modell):



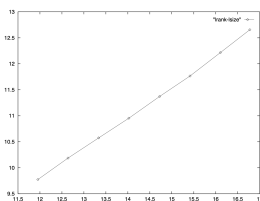
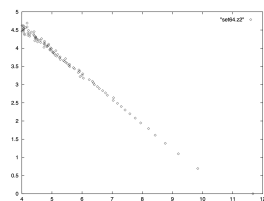
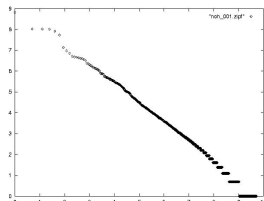
- Mi kell ehhez? Üzenetek egy (véges vagy végtelen) $\{m_1, m_2, \dots\}$ halmazára
- (Ha ismerjük, akkor jól jön egy N_{ij} mátrix ami megmondja, hogy az m_i bemenetnél milyen valószínűséggel jön ki m_j . Mi elsősorban a zajtalan ($M = I$) esettel fogunk foglalkozni.)
- Legyen az m_i valószínűsége p_i , ekkor $H = -\sum_i p_i \log p_i$ adja meg a modell információ-tartalmát (a csatorna kapacitását)
- Ha a log 2 alapú, akkor *bitekben* ha e alapú akkor *natokban*

A MONDATOK INFORMÁCIÓTARTALMA

- **Két részből áll:** a szavak által közvetített információból és a nyelvtani szerkezet által közvetített információból
- *Brutus megölte Cézár-t* és *Brutus-t megölte Cézár* ugyanazokat a szavakat (és ragokat) tartalmazza, és a sorrend nem számít (v.ö. *Cézár-t megölte Brutus*) de mást jelentenek: a különbséget az adja, hogy a tárgyeset *-t* ragja melyik személynéven jelenik meg.
- **Az első rész** Egy szó átlagosan $-\sum_i p_i \log p_i$ információt közöl. Ha jól ismerjük a szavak valószínűségét, akkor ez direktbe számolható. De mivel csak a gyakori szavak valószínűségét ismerjük jól, az összeget kettévágjuk a k -adik szónál: 1-től k -ig direktbe mérjük, onnantól *Zipf törvényével* számoljuk.

KIS KITÉRŐ: ZIPF TÖRVÉNYE

- Korpusz “szövegtest” sok, valamely kritérium által összeválogatott szöveg. Lehet zárt (pl. Shakespeare műveinek korpusza) vagy nyitott (WSJ korpusz – minden nap bővül)
- A korpuszra *mindig* mint egy eloszlásból vett mintára gondolunk, zárt korpusznál is (még előkerülhet egy ismeretlen munka).
- Ha megszámloljuk egy szó hányszor fordul elő, és a gyakoriság szerint csökkenő sorrendbe állítjuk a szavakat, azt mondjuk, hogy a leggyakoribb szó **rangja** 1, a következőé 2, stb.
- Ha a gyakoriságok logaritmusát a növekvő rangok logaritmusának függvényében ábrázoljuk, ilyen ábrát nyerünk:



ZIPF, ZIPF 2, HERDAN

- Zipf: $\log F(r) = C - B \log r$ azaz $F(r) = e^C / r^B$ ahol $e^C = N / \zeta(r)$ mivel $\sum F(r) = N$
- Zipf 2 “number frequency law”: $V(i, N) = mV(N) / i^{\rho+1}$
 $m = 1 / \zeta(\rho + 1), D = B / (1 + B)$
- Herdan's Law: $V(N) = N^{1/B}$
- B , a “Zipf konstans” jobban számolható Zipf második törvényéből mint az elsőből, és legjobban Herdan törvényéből számolható.
- A Zipf konstans függ a nyelvtől, olyan 1.1 és 1.5 között szokott lenni
- Aki akarja, megpróbálhatja kimérni

A SZÓENTRÓPIA BECSLÉSE

$$1 - P_k = C_k \sum_{r=k+1}^{N^{1/B}} r^{-B} \approx C_k \int_k^{N^{1/B}} x^{-B} dx = \frac{C_k}{(1-B)} [N^{\frac{1-B}{B}} - k^{1-B}]$$

$$C_k \approx (1 - P_k)(B - 1)k^{B-1}$$

$$H = - \sum_{r=1}^k p_r \log_2(p_r) - \sum_{r=k+1}^{N^{1/B}} p_r \log_2(p_r)$$

$$\frac{C_k B}{\log(2)} \int_k^{N^{1/B}} \log(x) x^{-B} dx - \frac{C_k \log(C_k)}{\log(2)} \int_k^{N^{1/B}} x^{-B} dx$$

$$H \approx H_k + \frac{1 - P_k}{\log(2)} (B/(B - 1) - \log(B - 1) + \log(k) - \log(1 - P_k))$$

esetünkben **12.6 bit/szó**.

A NYELVTANI ENTRÓPIA BECSLÉSE

Ez a második rész a mondatok entrópiabecslésében

Egy n hosszú mondatnak C_{n-1} elemzése lehetséges (bináris fák n pont felett). Rekurzió: $C_{n+1} = \sum_0^n C_i C_{n-i}$

A $C(x) = \sum C_k x^k$ generátorfüggvényre tehát teljesül a

$$xC(x)^2 = C(x) - 1$$

függvényegyenlet, amiből

$$C(x) = \frac{1 - \sqrt{1 - 4x}}{2x}$$

Ennek nyilván $1/4$ -nél van az első szingularitása, tehát $C_k \sim 4^k$ azaz a nyelvtani entrópia **2 bit/szó**.

MI EBBŐL A TANULSÁG?

- A grammatikai információ az egész információ mintegy 10-15%-a. (Az érzelmi információ, hangsúlyok, intonáció 1-2%).
- A hagyományos logikai szemantika kizárólag a grammatikai információval foglalkozik: Every man loves a woman
- $\forall \text{ man}(x) \exists \text{ woman}(y) \text{ Love}(x,y)$
- $\exists \text{ woman}(y) \forall \text{ man}(x) \text{ Love}(x,y)$
- A szójelentés “lexikai szemantika” megmaradt a szótárkészítőknek, mesterséges intelligencia/tudásreprezentáció (AI/KR) kutatóinak
- Mi a szótárkészítővel kezdjük, mert ők régóta csinálják (tehát sok a munkatapasztalat) és a végtermékek bizonyítottan hasznosak

bürtön fn 4C8 ◊

1. (*átv is*) 'szabadságvesztésre ítélt v. előzetes letartóztatásban levő **személyek őrzésére szolgáló büntetés-végrehajtási intézmény (épülete)**, ill. foglyok őrzésére, fogva tartására szolgáló egyéb épület, helyiség' ◊ *Börtön* Tömlöttz (1779 Baróti Szabó Dávid 7021009, 205) | *Börtönének* zárja nyílik: Ágnes a törvény előtt Megáll szépen, ahogy illik (1853 Arany János 8014059, 213) | *Hamlet*: Dánia *börtön*. Rosencrantz: Úgy az egész világ is az. *Hamlet*: De még milyen! mennyi rekesz, őrhely és dutyi van benne! (1867 Arany János ford.–Shakespeare 8014006, 133) | [A tanítóját megkéselő tanítvány] a *börtönben* várja büntetését (1894 Hazánk 8634001, 6) | kiszabadul a *börtönből* (1897 Eötvös Károly 8127001, 43) | méregkeverés miatt már *börtönben* is ült (1915 Kosztolányi Dezső 9359339, 60) | *fiatalkorúak bürtönébe* (1986 Solt Ottília 2003008, 378) | [Bernben a] *börtönök*, kantonális intézmények, a posta az állam fennhatósága alatt áll (1991 Vajna János 2003044, 30).

2. (*Jog is*) '**börtönbüntetés**' ◊ sebelés [= zsebmetszés] miatt *börtönre* ítéltetett (1841 Ormós Zsigmond 8341002, 196) | elítélték... Miklós.: Mire? *Liszka*.: Életfogytig való *börtönre* (1875 Tóth Ede 8486002, 108) | *Börtön*. A magy. btkvben a szabadságvesztés büntetéseknek egyik a fegyháznál enyhébb neme (1893 PallasLex. CD02) | A nyolcévi *börtön* után úgy tengette az életét, mint a bűnösök szokták: kilökte mindenhonnan, megint a bűneihez fordult (1931 Gelléri Andor Endre 9179008, 108) | mit tehetnének, hogy *börtön* nélkül ússza meg a galibát? (1985 Sütő András 9620021, 11).

3. (*irod, átv is*) '**tömlöcre emlékeztető szűk hely, amelybe vki, vmi be van szorítva, amely elzár, fogva tart vkit, vmit**' ◊ [nappal] a' Féreg-

-is úzi *Dolgait*; és kefkény *börtönbe* tekefzkedik [ti. bebábozódik] (1794 Baróti Szabó Dávid ford.–Vanière 7021011, 175) | a' szűk és tekervényes utcák' mély *börtönéből* (1845 Irinyi József 8200002, 115) | *Én* maradok: magam számára *börtön*, mert én vagyok az alany és a tárgy, jaj én vagyok az ómega s az alfa (1909 Babits Mihály CD01) | előborzonganak törékeny *börtönükből* [ti. az üvegből] az illanó eszenciáik (1937 Szabó Lőrinc 9629243, 69) | Ki húzta fel [a szívet], hogy rendületlen kopogjon *bordám bürtönében* (1980 Tollas Tibor 9818059, 236).

3a. 'az a hely(zet), körülmény, állapot, amelyben vki nem érzi magát szabadnak, amelyben nincs v. korlátozva van szabadsága' ◊ *Soká leszünk a' béke' bürtönében* (1826 Vörösmarty Mihály 8524395, 179) | *Egész gyermekségem egy bürtön* volt (1875 Arany János ford.–Szollogub C0662, 115) | ebből a bűvös körből, a pillanat *bürtönéből* nincs, nincs menekvés (1937 Karinthy Frigyes 9309002, 33) | „Nem emlékszem, én illet sosem akartam” – állítja, amikor a felettes én által *mintegy bürtönbe zárt* (tudattalanná tett) *gondolatról* [esik szó] (1986 F. Várkonyi Zsuzsa 1044008, 132).

Ö: állam~, pince~, vár~.

Fr: adós, adóssági.

ÖU: adós~, bástyá~, csillag~.

ÖE: ~ajtó, ~élet, ~épület, ~év, ~fogság, ~helyiség, ~intézet, ~kapu, ~kórház, ~lakó, ~reform, ~sziget, ~szoba, ~szolga.

Sz: bürtönész, bürtönöl, bürtönös.

Vö. CzF.; ÉrtSz.; TESz.; ÉKsz.; SzT.; ÚMTsz.

technically (the ~ feathers of his breast — Audrey A. Brown)

1 **prison** \ 'priz^ən\ *n* -s *often attrib* [ME *prison*, *prison*, *prison*, fr. OF *prison*, *prison*, fr. L *prehension*-, *prehensio* act of seizing, fr. *prehensus* (past part. of *prehendere* to seize, grasp) + *-ion*-, *-io* -ion — more at PREHENSILE] **1** : a place or condition of confinement or restraint (as of a person) : IMPRISONMENT (put in ~) (<~ seldom cures the criminal) **2** : a building or other place for the safe custody or confinement of criminals or others (as formerly debtors) committed by lawful authority; *often* : an institution for the imprisonment of persons convicted of major crimes or felonies : a penitentiary as distinguished from a reformatory, local jail, or detention home

2 **prison** \ '\ *vt* -ED/-ING/-S [ME *prisonen*, fr. *prison*, *n.*] *chiefly dial* : to put or keep in restraint : IMPRISON, CONFINE

börtön [-ök, -t, -e] *n*, 1. [*hely*] prison; gáol, jai, penitentiary (US), [*várban*] dungeon, [*dutyi*] clink ♦, quod ♦, hoosegow (US ♦); ~be juttat land sy in prison; ~be zár/vet Imprison, put/throw/shut/cast sy in prison, commit to prison, have sy sent to prison, send sy to prison, incarcerate (*ref*); ~be vetették be sent to prison, be/lie in prison; ~be vetés/zárás imprisoning, incarceration; ~ben ül/van be/lie in prison, be sent to prison, be in the lock-up (*fam*), serve a sentence of imprisonment, do time (*fam*), do one's stretch (*fam*); ~ben fogja végezni he will end up in prison; ~ből kiszabadul leave prison, be released from prison; megszökik a ~ből break (out of) prison, escape from prison; adósok ~e debtors' prison, [*ideiglenes*] sponging house (*obs*) 2. [*büntetés*] imprisonment, confinement, prison, penal internment; életfogytiglani ~ life imprisonment; öt évi ~ five years' imprisonment/confinement; ~nel büntet sentence sy to imprisonment, imprison

EGY KIS NYELVTAN

A híd leszakadt a teherautó súlya alatt

The bridge collapsed under the weight of the truck

Mi minden van ebben? Alany, állítmány, birtokos szerkezet, adpozíciós szerkezet, időjel, térbeli viszonyok, ágens? páciens? határozó?