Luonnollisen kielen tilastollinen käsittely

T-61.281 (3 ov) L

Luento 2, 22.1.2003

Luennot: **Timo Honkela**
Laskuharjoitukset: **Vesa Siivola**

Luentokalvot: Krista Lagus ja Timo Honkela

0.5 Laskuharjoitukset

Laskuharjoituksien ajaksi valittiin viime luennolla kolmesta vaihtoehdosta tiistai 16-18. Ensimmäiset laskuharjoitukset pidetään ti 28.1. paikkana sali Y313 TKK:n päärakennuksella.

Harjoitukset pitää DI Vesa Siivola (mailto:vesa.siivola@hut.fi).

Tehtävät laitetaan edellisenä perjantaina nähtäville osoitteeseen http://www.cis.hut.fi/Opinnot/T-61.281/laskarit.html.

3. MATEMAATTISIA PERUSTEITA

3.1 Todennäköisyyslasku

Peruskäsitteitä

Todennäköisyysavaruus (*probability space*):
Tapahtuma-avaruus Ω — diskreetti tai jatkuva
Todennäköisyysfunktio P
Kaikilla tapahtuma-avaruuden pisteillä A on todennäköisyys: $0 \leq P(A) \leq 1$

Todennäköisyysmassa koko avaruudessa on $\sum_A P(A) = 1$

Esimerkki 1

Jos tasapainoista kolikkoa heitetään 3 kertaa, mikä on todennäköisyys että saadaan 2 kruunaa?

Mahdolliset heittosarjat Ω: { HHH, HHT, HTH, HTT, THH, THT, TTH, TTT }

Heittosarjat joissa 2 kruunaa: A={ HHT, HTH, THH}

Oletetaan tasajakauma: jokainen heittosarja yhtä todennäköinen, $P = 1/8$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{8}$$

3.2 Ehdollinen todennäköisyys

A= asiaintila jonka todennäköisyyden haluamme selvittää
B= meillä oleva ennakkotieto tilanteesta, ts. tähän asti tapahtunutta

Ehdollinen todennäköisyys, A:n todennäköisyys ehdolla B:

$$P(A|B) = \frac{P(A,B)}{P(B)} \tag{1}$$

Palataan esimerkkiin 1: Oletetaan että on jo heitetty kolikkoa kerran ja saatu kruuna. Mikä nyt on todennäköisyys että saadaan 2 kruunaa kolmen heiton sarjassa?

Alunperin mahdolliset heittosarjat: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

Prioritiedon B perusteella enää seuraavat sarjat mahdollisia: { HHH, HHT, HTH, HTT }

$P(A|B) = 1/2$

3.3 Riippumattomuus

Kaksi tapahtumaa on tilastollisesti riippumattomia, jos niiden yhteinen todennäköisyys on sama kuin niiden erikseen tarkasteltujen todennäköisyyksien tulo:

$$P(A, B) = P(A)P(B) \tag{2}$$

Sama ilmaistuna toisin: se että saamme lisätiedon B ei vaikuta käsitykseemme A:n todennäköisyydestä, eli:

$P(A) = P(A|B)$

Tämä voidaan johtaa hyödyntäen em. ehdollisen todennäköisyyden kaavaa:

$$P(A|B) = \frac{P(A, B)}{P(B)} \tag{3}$$

Kausaalisuudesta ja riippuvudesta

Huom: tilastollinen riippuvuus $\neq$ kausaalinen riippuvuus!

Esim. jäätelön syönnin ja hukkumiskuolemien välillä voisi olla havaittavissa tilastollinen riippuvuus:

P('henkilö X hukkuu tänään','henkilö X on syönyt tänään jäätelöä') $>$
P('henkilö X hukkuu tänään')P('henkilö X on syönyt tänään jäätelöä')

Todennäköisyydet voisivat olla:

P('hukkuu tänään') $= 0.001$
P('hukkuu tänään'|'syönyt tänään jäätelöä') $= 0.002$
P('syönyt tänään jäätelöä') $= 0.2$

Ehdollinen riippumattomuus

$$P(A, B|C) = P(A|C)P(B|C) \tag{4}$$

A ja B ovat riippumattomia ehdolla C mikäli on niin että jos jo tiedämme C:n, tieto A:sta ei anna mitään lisätietoa B:stä (ja päinvastoin).

Edellisessä esimerkissä yhteinen kausaalinen tekijä on ehkä lämmin kesäsää:

P('hukkuu tänään','syönyt tänään jäätelöä'|'tänään lämmin ilma') =
P('hukkuu tänään'|'tänään lämmin ilma')P('syönyt tänään jäätelöä'|'tänään lämmin ilma')

Todennäköisyydet voisivat olla:

P('hukkuu tänään','syönyt tänään jäätelöä'|'tänään lämmin ilma') = 0.002
P('hukkuu tänään'|'tänään lämmin ilma') = 0.002
P('syönyt tänään jäätelöä'|'tänään lämmin ilma') = 1.0

3.4 Bayesin kaava

Paljon käytetty Bayesin kaava perustuu ajatukseen siitä, että koska kahden tapahtuman yhdessä esiintymisessä ei ole kyse kausaalisesta riippuvuudesta, tapahtumien järjestystä voidaan vaihtaa:

$$P(A, B) = P(B)P(A|B) = P(A)P(B|A) \tag{5}$$

Eli $P(B|A)$ voidaan laskea $P(A|B)$:n avulla.

$$P(B|A) = \frac{P(A, B)}{P(A)} = \frac{P(B)P(A|B)}{P(A)} \tag{6}$$

Thomas Bayes 1702-1761

T. Bayes. An Essay Towards Solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London, 53, pp. 370-418, 1763.

"I now send you an essay which I have found among the papers of our deceased friend Mr Bayes, and which, in my opinion, has great merit..."

Todennäköisimmän tapahtuman määrittely

Jos A = lähtötilanne, joka ei muutu (esim. jo tapahtuneet asiat), ja haluamme ainoastaan tietää, mikä tulevista tapahtumista B on todennäköisin, $P(A)$ on normalisointitekijä joka voidaan jättää huomiotta:

$$\arg\max_B P(B|A) = \arg\max_B \frac{P(B)P(A|B)}{P(A)} = \arg\max_B P(B)P(A|B) \quad (7)$$

Useampia kuin yksi ehto Bayesin kaavassa

$P(A)$ voidaan myös laskea useamman ehdon yhdistelmänä:

$P(A) = \sum_i P(A|B_i)P(B_i)$

Kannattaa huomata, että kaikille i $\neq$ j: $B_i \cap B_j = \emptyset$

3.5 Satunnaismuuttuja

Periaate: satunnaismuuttuja on se asia, josta ollaan kiinnostuneita, ja joka kussakin kokeessa saa jonkin arvon.

- Jatkuva-arvoinen satunnaismuuttuja: $X : \Omega => \mathbb{R}^n$, jossa $\mathbb{R}$ on reaalilukujen joukko ja n on avaruuden dimensio. Jos $n > 1$ puhutaan myös satunnaisvektorista.

- Diskreetti satunnaismuutuja: $X : \Omega => S$, jossa S on numeroituva $\mathbb{R}$:n osajoukko.

- Indikaattorimuuttuja: $X : \Omega => 0, 1 (Bernoulli - jakautunut)$.

Todennäköisyysjakauma *probability mass function pmf* $p(x)$ kertoo miten todennäköisyysmassa jakautuu satunnaismuuttujan eri arvojen kesken. Jakauman massa aina $= 1$ (muussa tapauksessa ei ole tn-jakauma).

3.6 Odotusarvo ja varianssi

Odotusarvolle $E(X) = \sum_x xp(x)$

(diskreetissä tapauksessa; jatkuvassa tapauksessa summan korvaa integraali)

Ts. odotusarvo on keskiarvo kussakin näytteessä (kokeessa) saadun satunnaismuuttujan arvon yli.

Varianssi kuvaa muuttujan arvon vaihtelua keskiarvon ympärillä:

$$\begin{aligned} Var(X) &= E((X - E(X))^2) \\ &= E(X^2) - E^2(X) \end{aligned}$$

(Keskihajonta lasketaan ottamalla varianssista neliöjuuri.)

3.7 Yhteisjakauma

Yhteistodennäköisyys

P(X,Y) = kahden tapahtuman tai väitteen yhteistodennäköisyys, ts. että molemmat toteutuvat (esim. X=jonain tiettynä ajanhetkenä kuultu sananmuoto on 'kuin' ja Y=samana ajanhetkenä kuullun sanan sanaluokka on substantiivi.)

Luetaan: X ja Y

Yhteistodennäköisyysjakauma

p(x,y) = kahden satunnaismuuttujan yhteisjakauma. Kuvaa x:n ja y:n kunkin arvokombinaation todennäköisyydet.

Kaavoja kootusti:

$$p(x,y) = p(X=x, Y=y) \quad \text{Yhteisjakauma} \quad (8)$$

$$p_X(x) = \sum_y p(x,y) \quad \text{Reunajakauma} \quad (9)$$

$$p(x,y) = p_X(x)p_Y(y) \quad \text{Riippumattomuus} \quad (10)$$

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} \quad \text{Ehdollinen jakauma} \quad (11)$$

$$p(x,y,z,w) = p(x)p(y|x)p(z|x,y)p(w|x,y,z) \quad \text{Ketjusääntö} \quad (12)$$

3.8 P:n laskeminen

- Yleisesti P on tuntematon, ja estimoitava datasta, tyypillisesti erilaisten tapahtumien frekvenssejä laskemalla.
- Koko todennäköisyysjakauman estimoinnin sijaan on mahdollista käyttää *parametrisia malleja* todennäköisyysjakaumille: tällöin estimoidaan vain jakauman parametrit.
- Bayeslaisessa estimoinnissa datan lisäksi huomioidaan prioritieto.

3.9 Esimerkki diskreetistä jakaumasta: Binomijakauma

Notaatio: Jakauma(satunnaismuuttuja; jakauman parametrit)

Satunnaismuuttujalla 2 mahdollista arvoa, onnistuu/ei, tai tarkasteltava ominaisuus (esim. tietty sana) joko on tai ei ole läsnä jossain tietyssä näytteessä (esim. lauseessa).

$p =$ Onnistumistodennäköisyys yksittäisessä kokeessa
$r =$ onnistumisten lukumäärä kun kokeita yhteensä n

$$b(r; n, p) = \frac{n!}{(n-r)!r!} p^r (1-p)^{n-r}, \text{jossa } 0 \leq r \leq n \tag{13}$$

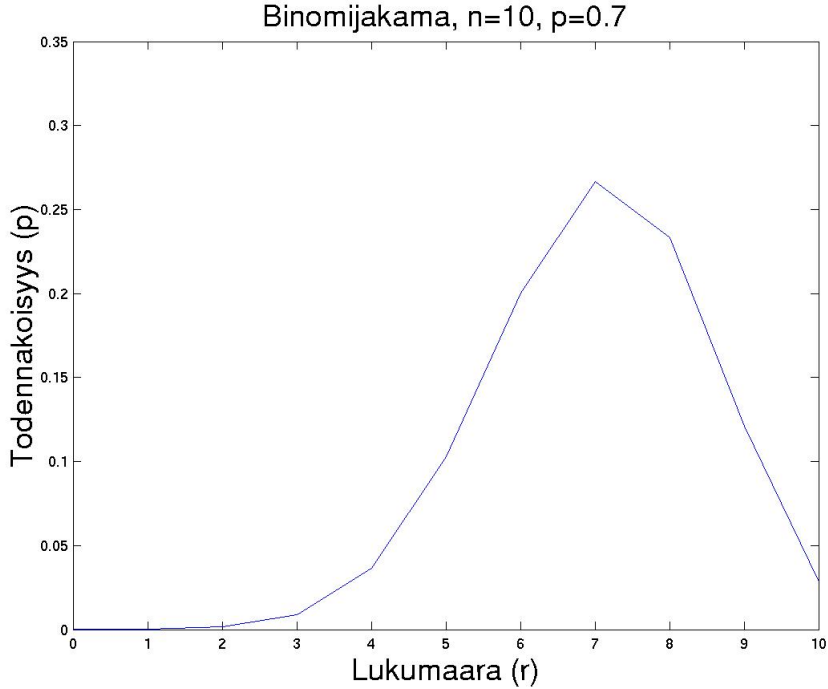
Odotusarvo: np, varianssi: $np(1-p)$

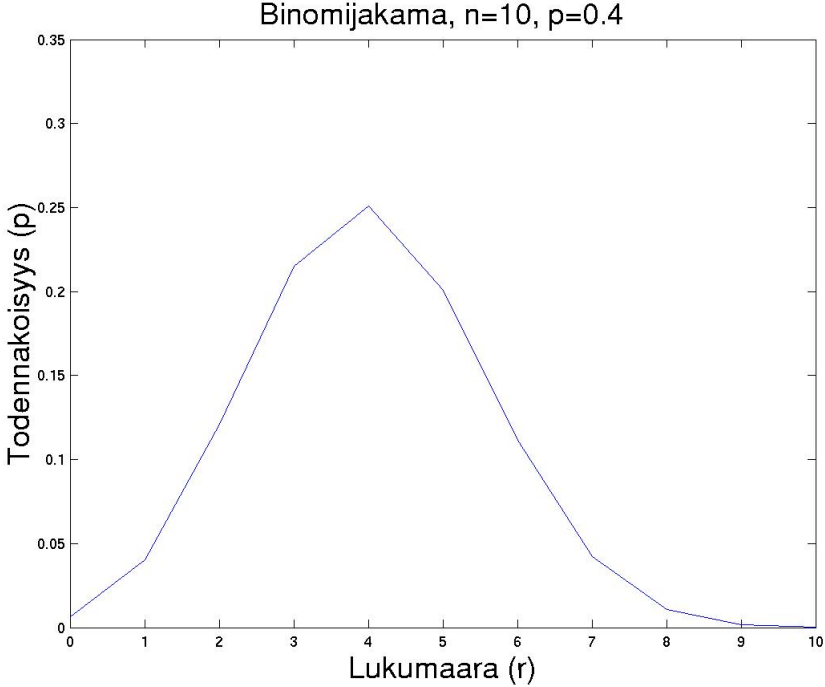
Oletus: riippumattomat kokeet. Kieleen sovellettaessa kuitenkin edellinen ja seuraava lause yleensä riippuvat toisistaan (samoin sanat), joten kokeet eivät ole todella riippumattomia.

Binomijakauman kuvaajaesimerkkejä

matlab-koodi:

```
> n = 10; p=0.7; for r = 0:n
> binomi(r+1) =
     factorial(n) / (factorial(n-r)*factorial(r))*
                    (p .^ r)*((1-p) .^ (n-r));
  end
> x = 1:n+1; plot(x-1,binomi(x))
```


Binomijakama, n=10, p=0.7
Todennakoisyys (p)
Lukumaara (r)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
7
8
9
10


Binomijakama, n=10, p=0.4
Todennakoisyys (p)
Lukumaara (r)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
7
8
9
10

Muita diskreettejä jakaumia

Multinomijakauma: Binomijakauman yleistys kun lopputuloksia voi olla useampi kuin kaksi.

Poisson-jakauma: Kiinteän kokoinen tapahtumaikkuna, jossa jakauma kuvaa tietyn, tutkittavan, asian tapahtumis- tai esiintymislukumäärän todennäköi
Satunnaismuuttuja x on siis tapahtumien lukumäärä tietyssä aikaikkunassa (tai jollakin matkalla, jollakin pinta-alalla tms.)

$$b(r; m) = \frac{m^r}{r!} e^{-m} \tag{14}$$

Parametri m on n:n ja p:n tulo $(n \times p)$. Jakauman varianssi ja keskiarvo on m

Poisson-jakauman sovellus

Tietyssä suuressa populaatiossa tiedetään aiemmin olleen 4 prosenttia erään kielen taitajia. Nykyistä tilannetta selvitettäessä valittiin populaatiosta satunnaisesti 200 henkilöä. Millä todennäköisyydellä 200 valitun joukossa on korkeintaan viisi k.o. kieltä osaavaa, jos populaatiossa osaajia on edelleen tuo 4 prosenttia.

Poisson-jakaumalla päästään kohtuulliseen likiarvoon.
Nyt $n \times p = 200 \times 0.04 = 8$.

$$P(X \leq 5) = \sum_{k=0}^{5} \frac{8^k}{k!} e^{-8} = 0.191 \tag{15}$$

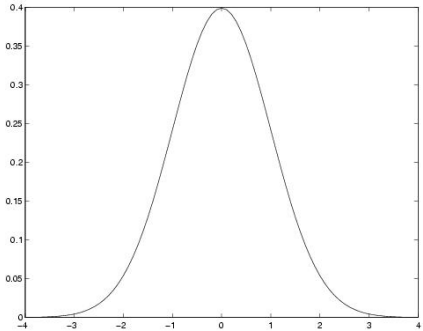
Normaalijakauma (gaussinen jakauma)

Määritelty, jos tunnetaan keskiarvo μ ja varianssi σ^2 :

$$N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x-\mu)^2/2\sigma^2}$$

Yleisesti: todennäköisyysjakauma voi olla mikä tahansa funktio jonka integraali $= 1$ välillä $[0, 1]$

Normaalijakauman kuvaaja



Vastaava matlab-koodi:

```
> x = -4:.1:4;
> y = 1/(sqrt(2*pi))*exp(-(x.^2)/2);
> plot(x, y);
```

3.10 Bayesläisestä tilastotieteestä

Tähän asti on tarkasteltu todennäköisyyttä *frekventistisestä* näkökulmasta.

Bayesläinen tulkinta: todennäköisyys kuvastaa *uskomuksen astetta*. Bayesläises mallinnuksessa myös prioritieto eli uskomukset *ennen* datan näkemistä ilmaistaan eksplisiittisesti.

Esimerkki 1: ainutkertaiset tapahtumat

Mikä on todennäköisyys sille että maailmankaikkeus loppuu huomenna?

Frekventisti: ei vastausta, koska koetta ei voi toistaa N kertaa.
Bayesläinen: subjektiivinen todennäköisyys (uskomus) on olemassa.

Esimerkki 2: taskussani olevien kolikoiden rahallinen arvo

Arvo on jokin täsmällinen luku, mutta tietoni siitä ovat vajavaiset. Uskomukseni: Arvo on varmasti positiivinen, ja lähes varmasti alle 20 euroa.

3.11 Bayesläinen päätösteoria

Optimaalinen tapa mallin (teorian) valintaan: valitaan malli (teoria), joka uskottavimmin selittää jonkin havaintojoukon.

Ts. maksimoidaan mallin todennäköisyys kun tunnetaan data ts. mallin *posterioritodennäköisyys*: $P(\text{Malli}|\text{data})$

Esimerkki 1: Mallin parametrien valinta

Kolikonheitto. Olkoon malli M_m joka sanoo $P(\text{kruuna}) = m, 0 \leq m \leq 1$. Olkoon s jokin heittojono jossa i kruunaa, j klaavaa.

$$P(s|M_m) = m^i(1-m)^j \tag{16}$$

Frekventistisestä näkökulmasta, valitaan malli joka maksimoi datan todennäköisyyden (*MLE, maximum likelihood estimate*):

$$\arg\max_m P(s|M_m) \tag{17}$$

Havainnot: 10 heittoa joista 8 kruunaa.
Frekventistinen lähestymistapa (MLE): $m = \frac{i}{i+j} = 0.8$

Bayesläinen lähestymistapa: kolikkoa tarkastelemalla näyttäisi siltä että kolikko on tasapainoinen, siis dataa katsomatta vaikuttaisi todennäköiseltä että $m = 1/2$ tai niillä main. Tämä uskomus voidaan liittää malliin *priorijakaumana mallien yli*.

Valitaan prioriuskomuksiamme kuvastava priorijakauma

Eräs sopiva priorijakauma olisi gaussinen jakauma jonka keskipiste (ja siis maksimi) on $1/2$:ssa. Valitaan kuitenkin prioriksi polynominen jakauma, jonka keskipiste (korkein kohta) $1/2$ ja pinta-ala 0 ja 1 välillä on 1:

$$p(M_m) = 6m(1-m)$$

Posterioritodennäköisyys Bayeslaisessa lähestymistavassa:

$$\begin{aligned} P(M_m|s) &= \frac{P(s|M_m)P(M_m)}{P(s)} \\ &= \frac{m^i(1-m)^j \times 6m(1-m)}{P(s)} \end{aligned}$$

jossa $P(s)$ on datan prioritodennäköisyys. Oletetaan, ettei se riipu mallista M_m joten voidaan jättää huomiotta mallia valittaessa.

Maksimoidaan osoittaja etsimällä derivaatan nollakohta m:n suhteen, kun $i = 8$ ja $j = 2$. Tämä on

$$\arg\max_m P(M_m|s) = \frac{3}{4} \tag{18}$$

Mallin estimointi on-line

Aloitetaan pelkällä priorimallilla, ja aina uuden havainnon tultua päivitetään malli posteriorimalliksi; ns. MAP (Maximum A Posteriori) -estimointi).

Taustaoletus: peräkkäiset havainnot ovat riippumattomia.

Esimerkki 2: Teorioiden tai malliperheiden vertailu

Havainnot: joukko aidan takaa kuultuja "kruuna" ja "klaava"- sanoja.

Malli/Teoria $M1(\theta)$: joku heittää yhtä kolikkoa, joka saattaa olla painotettu, ja mallin vapaa parametri θ on painotuksen voimakkuus.

Malli/teoria $M2$: joku heittää kahta tasapainoista kolikkoa, ja sanoo "kruuna" jos molemmat kolikot ovat kruunia, ja "klaava" muuten. Mallin $M2$ mukaan heittojonon, jossa on i kruunaa ja j klaavaa todennäköisyys on siis:

$$P(\text{data}|M2) = (\frac{3}{4})^i(\frac{1}{4})^j$$

Tehdään oletus: molemmat teoriat/mallit yhtä todennäköisiä *a priori* (ts. ennen kuin on saatu yhtään havaintoa): $P(M1) = P(M2) = 0.5$

Bayesin kaavasta:

$$P(M1|data) = \frac{P(data|M1)P(M1)}{P(data)} P(M2|data) = \frac{P(data|M2)P(M2)}{P(data)}$$

Halutaan selvittää kumpi malleista on uskottavampi. Lasketaan niiden uskottavuuksien välinen suhde:

$$\frac{P(M1|data)}{P(M2|data)} = \frac{P(data|M1)P(M1)}{P(data|M2)P(M2)}$$

Jos suhdeluku on > 1, valitaan malli $M1$, jos < 1, malli $M2$

(Vastaukset eri heittosarjoilla: taulukko 2.1 kirjan sivulla 58)

3.12 Shannonin informaatioteoria

- Claude Shannon, 1948 ("A Mathematical Theory of Communication")
- Tavoitteena maksimoida informaation siirtonopeus kohinaisella kommunikaatikanavalla
- Teoreettinen maksimi datan pakkaamiselle on entropia (H)
- Kanavan kapasiteetti C: jos kapasiteettia ei ylitetä, virheiden todennäköisyys saadaan niin alhaiseksi kuin halutaan.
- Nykyiset tiedonpakkausmenetelmät hyödyntävät näitä teoreettisia tuloksia.

Entropia

Olkoon $p(x)$ satunnaismuuttujan X jakauma diskreetin symbolijoukon (aakkoston) A yli:

$$p(x) = P(X = x), x \in A$$

$$H(p) = H(X) = -\sum_{x \in A} p(x) \log_2 p(x) \tag{19}$$

(Määritellään $0 \log 0 = 0$).

Entropia ilmaistaan tavallisesti biteissä (kaksikantainen logaritmi), mutta muunkantaiset logaritmit yhtä lailla ok.

Jos symbolijoukko on tasajakautunut, entropia on maksimissaan.

Esimerkki: 8-sivuisen nopan heittäminen, kommunikoitava yksittäisen heiton tulos.

$$\begin{aligned} H(X) &= -\sum_{i=1}^{8} p(i) \log p(i) = -\sum_{i=1}^{8} \tfrac{1}{8} \log \tfrac{1}{8} \\ &= -\log \tfrac{1}{8} = log 8 = 3\text{bittiä} \end{aligned}$$

Pätee yleisesti: Jos viestin todennäköisyys on $p(i)$, sen optimaalinen koodinpituus on $-\log p(i)$ bittiä.

Vaihtoehtoinen kirjoitustapa entropian kaavalle:

$$\begin{aligned} H(X) &= -\sum_{x \in A} p(x) \log p(x) = \sum_{x \in A} p(x) \log \tfrac{1}{p(x)} \\ &= E(\log \tfrac{1}{p(x)}) \end{aligned}$$

ts. entropia = optimaalisen koodinpituuden odotusarvo, eli montako bittiä keskimäärin on käytettävä yhden viestin välittämiseen.

Yhteisentropia ja ehdollinen entropia

Kahden muuttujan X ja Y (aakkostot A ja B) yhteisentropia, eli paljonko informaatiota keskimäärin tarvitaan kummankin arvon kommunikointiin:

$$H(X,Y) = -\sum_{x \in A} \sum_{y \in B} p(x,y) \log p(X,Y) \tag{20}$$

Ehdollinen entropia: Jos X on jo kommunikoitu, paljonko lisäinformaatiota keskimäärin tarvitaan Y:n kommunikoimiseen:

$$\begin{aligned} H(Y|X) &= \textstyle\sum_{x \in A} p(x) H(Y|X=x) && (21) \\ &= -\textstyle\sum_{x \in A} \sum_{y \in B} p(x,y) \log p(y|x) && (22) \end{aligned}$$

Entropian ketjusääntö: $H(X,Y) = H(X) + H(Y|X)$

Erikoistapaus: Jos muuttujat riippumattomia toisistaan, kumpikin voidaan kommunikoida erikseen ja laskea koodinpituudet yhteen:
$H(X,Y) = H(X) + H(Y)$

Vrt. todennäköisyyksien ketjusääntö $P(X,Y) = P(X)P(Y|X)$
ja riipumattomille muuttujille $P(X.Y) = P(X)P(Y)$.

Yhteisinformaatio (Mutual Information, MI)

Yhteisinformaatio I muuttujien X ja Y välillä on

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{23}$$

Kohinainen kanava-malli

Binäärinen kommunikointikanava, lähetetään 1 tai 0.
$p =$ todennäköisyys jolla kanavalla lähetetty bitti kääntyy päinvastaiseksi.

Kanavan kapasiteetti C on tällöin:

$$C = \max_{p(X)} I(X;Y) = 1 - H(p) \tag{24}$$

(kaavan johto kirjassa)

Relevanssi kielen mallintamisessa

Dekoodausongelmina voidaan tarkastella esimerkiksi

- konekäännöstä
- merkkien tunnistusta (OCR)
- puheentunnistusta

3.13 Minimum Description Length (MDL) -periaate

- Lähestymistapa mallin valintaan
- Rissanen et. al.
- Tavoite: pyritään löytämään datalle sellainen koodi että koko datajoukon koodauspituus minimoituu
- Koodinpituus = mallin kuvauspituus + datan kuvauspituus mallin avulla koodattuna + virheiden koodauspituus
- Koodinpituutta (todellista tai laskennallista) käytetään kustannusfunktiona mallia optimoitaessa
- Teoreettinen alaraja koodinpituudelle: entropia
- Suora yhteys myös Bayesläiseen mallinnukseen

4. Yleisen kielitieteen perustietoja

Itse luettavaa: kirjan luku 3, WWW-sivun linkit

4.1 Kielellisen analyysin eri tasoista

Käsiteltäviä kielellisiä yksiköitä

foneemi, morfeemi, sananmuoto, lekseemi, käsite, lause, virke, kappale, dokumentti, korpus

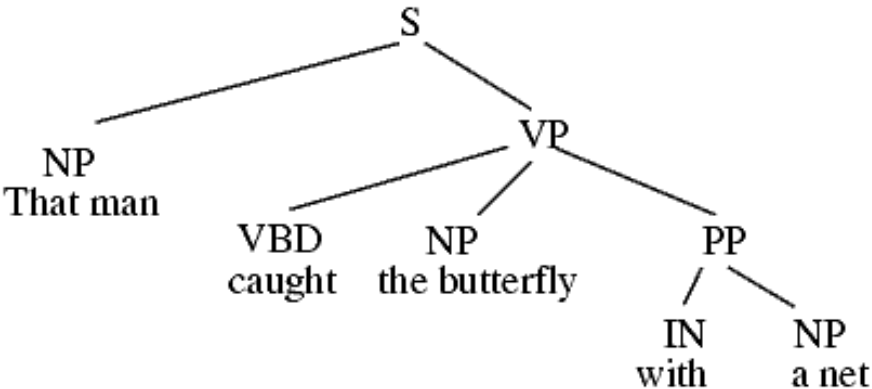
Tiedon lajeja, eri yksiköiden tasoilla

foneettinen ja fonologinen, morfologinen, syntaktinen, semanttinen, pragmaattinen, diskurssitieto, maailmantieto

Esimerkki syntaktisesta analyysista

Käsitteitä: Sanakategoriat, Lauserakennekielioppi, Dependenssikielioppi

Lauserakennekieliopin jäsennyspuu



Esimerkki morfosyntaktisesta analyysistä

Tuotettu Conexorin FDG:llä
FDG=Functional Dependency Grammar

Kääpiösilkkiapina	kääpiö#silkki#apina	&NH N SG NOM
on	olla	&+MV V ACT IND PRES SG3
niin	niin	&ADV ADV &AD> ADV
pieni	pieniä	&+MV V ACT IND PAST SG3
,	,	PUNCT
että	että	&CS CS
se	se	&NH PRON SG NOM
mahtuu	mahtua	&+MV V ACT IND PRES SG3
kämmenelle	kämmen	&NH N SG ALL
.	.	PUNCT