

SZEMANTIKA, 7. ELŐADÁS

Kornai András

BME 2020/10/21

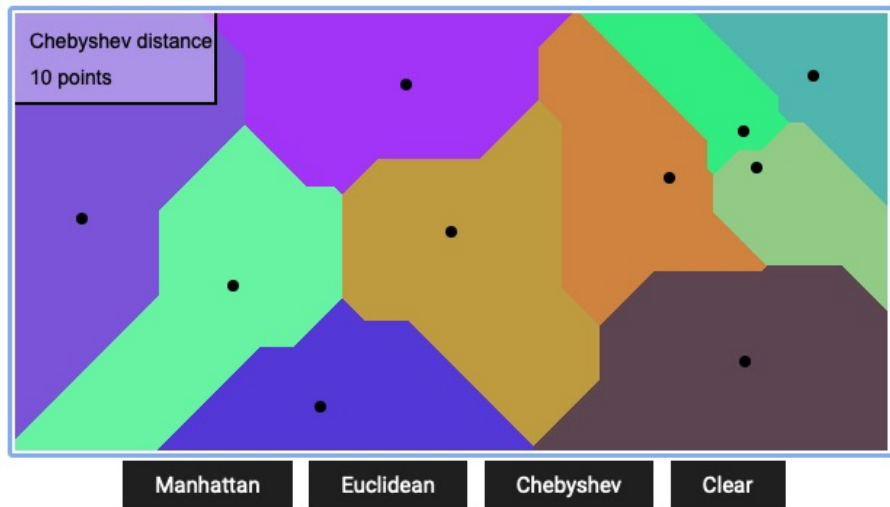
PROJEKTEK

- A tulajdonnevek és genusaik (Tauber Boglárka)
- Generikusok
- Alapértelmezések (defaults) (Eper Miklós)
- Határozók
- Sokklauzús definíciók (Koncz Máté)
- Más linkerek mint =AGT, =PAT (Havas Tamás)
- Adatok begyűjtése (pl.)

ENGEDÉLYKÉRÉS HF

- =AGT engedélyt kér X-től 'valaki mástól' Y-ra
- *kér valaki valakitől valamit* request =AGT express/2757 [=AGT want [=AGT HAS =PAT]] ebben nincs benne hogy kitől kéri, ugyanúgy, ahogy az sincs benne, hogy mennyi időre kéri, milyen célból, stb.
- De ha kifejezzük, akkor az a dolog forrása (ablatívusz) nem pedig (mondjuk) a kért dolog birtokosa (genitívusz) vagy helye (lokatívusz) *kérem a Peti játékát, kérem a fogót a fiókból.*
- Az adat megoszlik *kér valamit valakitől* és *kér valakit valamire* közt. Ez igen gyakran szemantikailag motiválatlan: miért *ra/re* kér? Semmi köze a *leteszi az asztalra, felnéz a toronyra* (lokatívuszi) alakokhoz
- Azt mondtuk, hogy mindennek ami nem levezethető annak a szótárban a helye, akkor tehát ennek is. Hagyományos nyelvészeti megoldás: **vonzatkeret**. Felsoroljuk az összes idioszinkratikus dolgot.

VORONOI DIAGRAMM



<http://yunzhishi.github.io/vorono.html>

DEFINÍCIÓ

Egy $V = \langle \mathcal{P}, P \rangle$ voronoid \mathbb{R}^n -beli politópok $\mathcal{P} = \{P_i\}$ olyan családja, melynél minden P_i belsejében van pontosan egy kitüntetett p_i pont

- 1 Voronoi diagrammokat használnak pszichológiai osztályozásnál (Gardenfors 2000), a voronoidok ennél általánosabbak, mert
- 2 nem várjuk el, hogy a p_i a P_i közepére essen
- 3 nem várjuk el, hogy a politópok határoló (hiper)síkjai egyenlő távolságra legyenek valamely két kitüntetett ponttól
- 4 nem várjuk el, hogy a P_i -k uniója lefedje az egész teret (teljes, pozitív mértékű régiók hiányozhatnak)

PAC TANULÁS + RITKASÁG

LINEARITÁS

Egy voronoid *lineáris* ha van hipersíkok egy olyan h_j halmaza, amely a politópok összes oldalát tartalmazza

PAC TANULÁS

Minden c konceptum megfelel egy π_c valószínűségi eloszlásnak \mathbb{R}^n felett

Egy konceptumnak vannak verbális deskriptorai 'hengeres, a tengelyében kanóccal, általában viaszból, elsősorban ünnepélyes alkalmakkor használatos világítóeszköz' de vannak nem-verbálisak is, pl. "a gyertya" vizuális képe, vagy az égő gyertya jellegzetes szaga
Két célunk van: egyrészt minden konceptumot bele akarunk foglalni egy P_i politópba úgy, hogy $\pi_c(P_i)$ kielégítőn közel legyen 1-hez, másrészt csökkenteni akarjuk a hipersík-rendszer méretét

BEÁGYAZÁS (STATIKUS)

Ha adott egy D szótár, a statikus beágyazás egy olyan \vec{v} függvény, ami minden $w \in D$ szóhoz egy $\vec{v}(w) \in \mathbb{R}^n$ vektort rendel

- Az ötletet először Schutze:1993 javasolta
- Az első ténylegesen működő implementáció Bengio:2003
- NLP “majdnem semmiből” POS, CHUNK, NER, mélyeset-cimkézés Collobert:2011
- Lineáris struktúrája van (király – királynő = férfi –nő) Mikolov:2013l
- De miért? Pennington:2014, Arora:2015, Gittens:2017

A BEÁGYAZÁSOK ALAPTULAJDONSÁGAI

1. Gyakoriság

$$\log(p(w)) = \frac{1}{2d} \|\vec{w}\|^2 - \log Z \pm o(1) \quad (1)$$

2. Együttes előfordulási becslés

$$\log p(w, w') = \frac{1}{2d} \|\vec{w} + \vec{w}'\|^2 - 2 \log Z \pm o(1) \quad (2)$$

3. PMI (pontenkénti kölcsönös információ)

$$\langle \vec{w}, \vec{w}' \rangle \sim \frac{\log p(w, w')}{\log p(w) \log p(w')} \quad (3)$$

4. Többértelmű szavak: általában mint az egyes szóértelmek gyakorisággal súlyozott összege. Újabban: kontextus által egyértelműsített *dinamikus* beágyazások (ELMO, NERT)

KOMMUNIKÁCIÓ VORONOIDOK

SEGÍTSÉGÉVEL

Ha a szóvektorokat tesszük meg a voronoidok kitüntetett pontjainak, már tudunk is kommunikálni

CIMKÉZÉS

A \vec{v} beágyazás inverze egy $l : \mathbb{R}^n \rightarrow D^*$ **cimkefüggvény** amely a politóp minden pontjához ugyanazt a szótári cimkét rendeli mint a kitüntetett pontjához.

Ha a beszélőnek és a hallgatónak hasonló voronoidja vannak, akkor egyszerű érzeteket és eszméket úgy lehet kommunikálni, hogy kimondjuk a cimkét: látok egy gyertyát, és azt mondom *gyertya*. Ez elegendő a hallgatónak ahhoz, hogy tudja melyik politóp volt a gondolataimban aktív, és ebből az én gondolataimról/érzéseimről megtudjon valamit. Ehhez nem kell, hogy a $P_{gyertya}$ politópjainak pont ugyanazok legyenek a határai, csak az, hogy a $\pi(\text{gyertya})$ valószínűségi tömeget nagyjából mindkettő lefedje.