Lexikální disambiguace a textual entailment

Silvie Cinková

UFAL, 25.1.2012

Předpoklady Word Sense Disambiguation

- Mluvčí jednoho jazyka sdílejí inventář navzájem se vylučujících „slova smyslů".

- Lexikografové tento inventář umějí popsat.

- Pro obyčejné lidi je snadné pro slovo k kontextu vybrat příhodný smysl z inventáře.

- Když se tuto disciplínu naučí stroj, zlepší se výkon NLP aplikací.

Praxe WSD

- anotace:
  - konflikt shody a sémantické granularity
  - WordNet: granulární, ale nespolehlivý
  - PropBank, OntoNotes: spolehlivější, ale málo informace
  - FrameNet: shodu nikdo neměří (???)
- aplikace:
  - špatný výkon WSD modulů, celkový výkon spíše zhoršují
  - disambiguace probíhá i v jiných modulech (koreference, NE...), samostatný WSD modul asi ani není potřeba

Zdroje pro WSD

- ## Word Sense Disambiguation
  - slovu v kontextu přiřadit z inventáře významů jediný, jehož definice nejlépe odpovídá
  - definice může být
    - sémantická (OntoNotes)
    - syntaktická (PropBank, Engvallex)

OntoNotes- Wordsenses

- actively search for something or someone, strive for something

Are you actively seeking jobs?

The authorities are seeking answers.

Legal advice should **be sought** before you take any further action.

We **sought out** the best guitarist in town for our band.

The police **seek out** new clues.

This software will **seek out** security holes in your system.

The candidates **seek election**.

Graffiti artists seek fame through vandalism in Los Angeles.

They **sought to reassure** the public.

I **seek to buy** several volumes for my book collection.

- go towards a place or a thing

Water seeks its own level.

seek

PropBank

- **Roleset id: seek.01 , *look for,* vncls: 35.6, framnet: Seeking_to_achieve , Seeking**
- **seek.01**: Member of VNcls ferret-35.6.
  Roles:
-     **Arg0**: *Agent / entity seeking* (vnrole: 35.6-agent)
  Arg1: *thing sought, attempted action* (vnrole: 35.6-theme)
  Arg2: *sought-from* (vnrole: 35.6-location)
  Example: transitive
-     Northeast is seeking approval of its possible purchase of PS of New Hampshire.
  Example: arg1 as action
-     A company is seeking *trace* to purchase a carrier.

  Example: with sought-from
-     Czechoslovakia is seeking $2 billion from Hungary.

Engvallex

ACT()[seek.01::0::260,seek.01::2::1] PAT()[seek.01::1::368]   ?ORIG()[seek.01::2::13,seek.01::1::4]
[seek-v.xml::seek.01::362]
(look for: transitive)
Northeast is seeking approval of its possible purchase of PS of New Hampshire.
Czechoslovakia is seeking $2 billion from Hungary.
A company is seeking *trace* to purchase a carrier. (v-u_nobody)
----

seek (**seeks** *3rd person present*) (**seeking** *present participle*) (**sought** *past tense & past participle*)

1 **verb** If you seek something such as a job or a place to live, you try to find one.
FORMAL They have had to seek work as labourers... *V n*
Four people who sought refuge in the Italian embassy have left voluntarily... *V n*
Candidates are urgently sought for the post of Conservative party chairman. *be V-ed for n*

2 **verb** When someone seeks something, they try to obtain it.
FORMAL The prosecutors have warned they will seek the death penalty... *V n*
Haemophiliacs are seeking compensation for being given contaminated blood. *V n*

3 **verb** If you seek someone's help or advice, you contact them in order to ask for it.
FORMAL Always seek professional legal advice before entering into any agreement... *V n*
The couple have sought help from marriage guidance counsellors. *V n from n*

4 **verb** If you seek to do something, you try to do it.
FORMAL He also denied that he would seek to annex the country... *V to-inf* **seek out** **phrasal verb** If you seek out someone or something or seek them out, you keep looking for them until you find them.
Now is the time for local companies to seek out business opportunities in Europe... *V P n (not pron)*
Ellen spent the day in the hills and sought me out when she returned. *V n P*

WSD vs. podobnost

- WSD pro člověka obtížné, nízká shoda
- vyšší shoda u posuzování podobnosti
  - Erk: slovu v kontextu přiřaď skóre podobnosti u každého slovníkového významu
  - Rumshisky: oskóruj větu, do jaké míry je v ní sloveso použito ve stejném významu jako ve vzorové

Další sémantické disciplíny

- Distribuční sémantické modely (Erk, 2010)
- Recognizing textual entailment (Zanzotto, 2009)

Ani jedna z disciplín nepředpokládá statický inventář významů – snaží se do synakt. a morfo. analýzy nějak dodat znalost světa:

- WordNet
- PropBank
- FrameNet
- OntoNotes

Ido Dagan, Dan Roth, Fabio Massimo Zanzotto (ACL tutorial 2007)

TEXTUAL ENTAILMENT

Applied Textual Entailment

- A directional relation between two text fragments:
  Text (t) and *Hypothesis (h):*

***t entails h* (*t$\Rightarrow$h*) if**

humans **reading *t* will infer that *h* is** most likely **true**

- **Operational (applied) definition:**
  - **Human gold standard - as in NLP applications**
  - **Assuming common background knowledge – which is indeed expected from applications**

Revised definition of an Old Problem: Sense Ambiguity

- Classical task definition - interpretation:
  Word Sense *Disambiguation*

- What is the RIGHT set of senses?
  - Any concrete set is problematic/subjective
  - ... but WSD forces you to choose one

- A lexical entailment perspective:
  - Instead of identifying an explicitly stipulated sense of a word occurrence ...
  - identify whether a word occurrence (i.e. its *implicit* sense) entails another word occurrence, in context
  - Dagan et al. (ACL-2006)

Some suggested research directions

- Knowledge acquisition
  - Unsupervised acquisition of linguistic and world knowledge from general corpora and web
  - Acquiring larger entailment corpora
  - Manual resources and knowledge engineering
- Inference
  - Principled framework for inference and fusion of information levels
  - *Are we happy with bags of features?*

Complementary Evaluation Modes

- **"Seek" mode:**
  - Input: h and corpus
  - Output: all entailing t's in corpus
  - Captures information seeking needs, but requires post-run annotation (TREC-style)
- **Entailment subtasks evaluations**
  - Lexical, lexical-syntactic, logical, alignment…
- ***Contribution to various applications***
  - QA – Harabagiu & Hickl, ACL-06;
    RE – Romano et al., EACL-06

Patterns – vhodný zdroj pro RTE?

- Vznikají jako shluky na základě vzájemné podobnosti.-> menší problém se shodou?
- Každý shluk frekventovaných konkordancí s taxonomickým popisem participantů dodává znalost světa *(co typicky dělají instituce, co se děje dokumentům?, co se dává k dispozici?, co se dostane do rozpaků?)*

Více přípustných hodnot místo slévání

- Implikatury shluků se navzájem sémanticky **odlišují**, ale **nevylučují**
- Často je nemožné určit jen jednu, která je centrální
- *Christians were thrown to the lions.*
  - *throw objects (propel by hand)*
  - *throw garbage (discard)*
  - *throw someone against soething, on the floor, etc.*
  - *throw objects to someone (pass)*
  - *throw corn to chickens*
  - *throw a corpse (treat a body badly, get rid of it)*