Introduction to Human Language Technologies
6. Word Sense Disambiguation

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona

Outline

Motivation

English word *dog* in a sentence, how should it be translated to Spanish?

NOUN

1. (animal)

a. **el perro (m)**, **la perra (f)**

My dog is a German Shepherd. — Mi perro es un pastor alemán.

2. (colloquial) (wicked person)

a. **el bribón (m)**, **la bribona (f)**

My coworker is a lazy dog; I'm always having to do his work. — Mi colega es un bribón perezoso; siempre le tengo que estar haciendo el trabajo.

b. **el canalla (m)**, **la canalla (f)** (colloquial)

That dog started cheating on his girlfriend almost as soon as they started going out. — Ese canalla le pegó cuernos a su novia prácticamente tan pronto empezaron a salir.

3. (pejorative) (unattractive woman)

TRANSITIVE VERB

4. (to follow)

a. **seguir**

The neighborhood bullies dogged him all the way to his house. — Los matones del vecindario lo siguieron el camino entero hasta llegar a su casa.

5. (to plague)

a. **perseguir**

He has been dogged by scandal his entire career. — El escándalo lo ha perseguido durante su carrera entera.

Source: `http://www.spanishdict.com`.

Lexical Ambiguity & Word Senses

sense	gloss from WordNet 1.5
age 1	the length of time something (or someone) has existed
age 2	a historic period

He was mad about stars at the age of nine .

WSD has been defined as AI-complete (Ide & Véronis, 1998);
such as the representation of world knowledge

Usefulness of WSD

- WSD is a potential intermediate task for many other NLP tasks
- WSD capabilities are involved in many applications:
  - Machine Translation
  - Information Retrieval
  - Semantic Parsing
  - Speech Synthesis and Recognition
  - Natural Language Understanding
  - Acquisition of Lexical Knowledge
  - Lexicography
- Unfortunately, this usefulness has still not been demonstrated

Resources

- Sense Definitions
  - Machine Readable Dictionaries
  - Bilingual Machine Readable Dictionaries
  - WordNets (large lexical databases)
- Corpora
  - Samples with only one word labeled for each sample
    - SemEval Lexical Sample Task (training/Test corpus)
    - mainly for supervised Machine Learning algorithms

  800004
  Mr Purves is tight-lipped about what happens then.
  He vexed rumour-mongers, who <tag '520051'>bet</> on a bid for Midlan sooner.
  800005
  Mr Jones loses his <tag '519914'>bet</>:1,000 people attended Cowley pools last year.

  - Samples with all words labeled
    - Semcor, SemEval All Words Task (Test corpus)
    - mainly for unsupervised algorithms
- Evaluation exercises: SensEval and SemEval

Outline

Types of WSD Algorithms

- Classification according to the information source:
  - **knowledge-based**: from a external knowledge source
    - Example: Lesk Algorithm available at NLTK, UKB (Graph-based model: Page Rank)
  - **supervised corpus-based**: examples with its correct sense
    - Examples: Naïve Bayes, kNN or SVM
  - **semisupervised corpus-based**: most of the examples with no sense information
    - Example: Yarowsky Algorithm (Boostrapping)
  - **unsupervised corpus-based**: examples with no sense information
    - Example: word embeddings + deep learning (to see in AHLT)

Lesk algorithm

Lesk algorithm
Disambiguates just one word within a context

$$\mathrm{Lesk}(w) = \underset{s_i \in S(\{w\})}{\mathrm{argmax}} \; \forall_{s_j \in S(C(w))} |\mathrm{Def}(s_i) \cap \mathrm{Def}(s_j)|$$

$S(X)$: set of senses for all lemmas in X

$C(w)$: set of lemmas in the context of word w.

$\mathrm{Def}(s)$: set of lemmas in the definition of sense s.

Lesk algorithm: example

Input: "pine cone"

PINE

1. kinds of evergreen tree with needle-shaped leaves
2. waste away through sorrow or illness

CONE

1. solid body which narrows to a point
2. something of this shape whether solid or hollow
3. fruit of certain evergreen trees

Solution

The best intersection is $Pine\#1 \cap Cone\#3 = 2$.
sense for "pine": $Pine\#1$
sense for "cone": $Cone\#3$

Lesk algorithm: simplification

Simplified Lesk algorithm

$$\mathrm{Lesk}(w) = \underset{s_i \in S(\{w\})}{\text{argmax}} \; |\mathrm{Def}(s_i) \cap \mathrm{C}(w)|$$

$S(X)$: set of senses for all lemmas in X

$\mathrm{C}(w)$: set of lemmas in the context of word w.

$\mathrm{Def}(s)$: set of lemmas in the definition of sense s.

In general, better performance that the general Lesk algorithm

Lesk algorithm: exercise

Given the sentence:

- `I went to the bank to deposit money.`

and the definitions of the two first senses of the word *bank*:

1. `sloping land (especially the slope beside a body of water)`
2. `a financial institution that accepts deposits and channels the money into lending activities`

apply simplified Lesk algorithm to find the most appropriate sense among them.

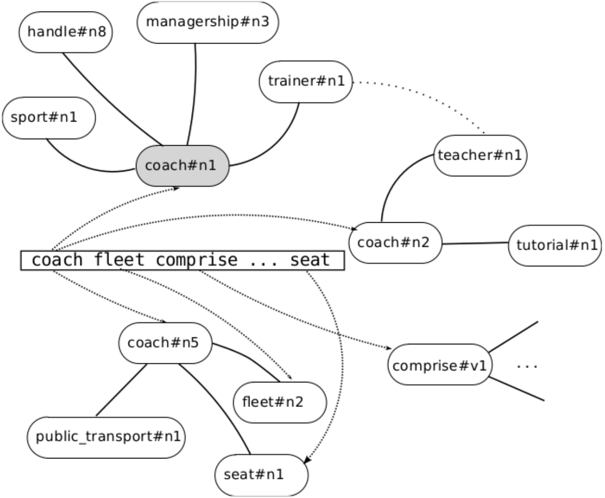
Lesk's algorithm: extensions

- Stopwords list
- Changing the similarity measure: Cosine
- Use examples of Wordnet Synsets
- Use the data of hypernyms and/or hyponyms
- Enrichment with WordNet (Adapted/Extended Lesk) (Banerjee and Pederson, 2002/2003)
- Enrichment with WordNet and Wikipedia (Enhanced Lesk) (Basile et al. 2014)

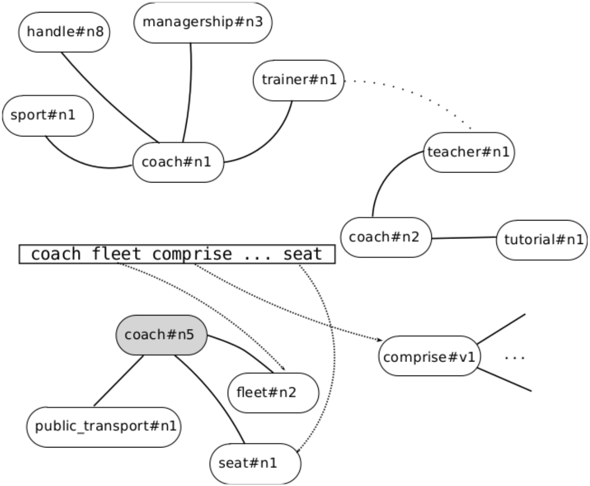
UKB algorithm (Agirre and Soroa, 2009)

- Disambiguates all words within a context at once
- WordNet is seen as a concept graph
  (each synset is a node and each lexical relation as a bidirectional edge)
- The context words (also target words) are included into the graph:
  - with directed edges to their senses
  - as source nodes injecting mass to the concept graph
- Page Rank algorithm is used to compute the weight of the nodes
  - General idea: weight each node by taking into account the number of nodes pointing to it and the weight of such nodes
  - UKB is a small variant avoiding cycles

UKB example


handle#n8
managership#n3
trainer#n1
sport#n1
coach#n1
teacher#n1
coach#n2
tutorial#n1
coach fleet comprise ... seat
coach#n5
comprise#v1
...
fleet#n2
public_transport#n1
seat#n1

UKB example



handle#n8
managership#n3
trainer#n1
sport#n1
coach#n1
teacher#n1
coach#n2
tutorial#n1
coach fleet comprise ... seat
coach#n5
comprise#v1
...
fleet#n2
public_transport#n1
seat#n1

Supervised Corpus-based

- WSD as a Machine Learning classification problem:
  - learn a different model to disambiguate each word
  - classes: senses of the word (Y vector)
  - examples: word occurrences in a sentence + correct sense representation with attributes such as:
    - local context (collocations, bigrams):
      i.e. word on the right is a verb
    - topic or broad-context (bag of words or vector space model):
      i.e. word "years" occurs in the sentence
    - syntactic features:
      i.e. its subject is çat"
    - domain information:
      i.e. the example is about *history*'
- 'knowledge acquisition bottleneck'

  the lack of widely available semantically tagged corpora, from which to construct really broad coverage WSD systems, and the high cost in building one

Exercise

We want the sentence below to be represented by local and topical features and be supply as example for a ML algorithm:

Example He was mad about stars at the age of nine .
age.01

+ PoS ('He', 'PRP'), ('was', 'VBD'), ('mad', 'JJ'), ('about', 'IN'), ('stars', 'NNS'), ('at', 'IN'), ('the', 'DT'), ('age', 'NN'), ('of', 'IN'), ('nine', 'CD'), ('.', '.')

1. Give the bag of open-class words of the left context.
2. Give the local features in a ± 2 word window of the word forms.
3. Give two other possible local or topical features