Word Sense Disambiguation

Farshad Borjalizade, Dr.BabaAli

NLP, Spring 2021, University of Tehran

Introduction

- **Word sense = distinct meaning of a word**

Same word, different senses

- **Polysemy :** Word have multiple senses
  - financial bank
  - blood bank
  - tree bank
- **Homonymy :** unrelated senses
  - May I come in?
  - Let's meet again in May?

Different word, same sense

- **Synonymy**

Part of speech ambiguity

- Joe won the first round
- Joe has a round toy

How big is the problem?

- **Most words in English have only one sense**
    - 62% in Longman's Dictionary of Contemporary English
    - 79% in WordNet

How big is the problem?

- **Most words in English have only one sense**
  - 62% in Longman's Dictionary of Contemporary English
  - 79% in WordNet
- **But the others tend to have several senses**
  - Average of 3.83 in LDOCE
  - Average of 2.96 in WordNet
- **Ambiguous words are more frequently used**
  - In the British National Corpus, 84% of instances have more than one sense
  - Some senses are more frequent than others

WordNet

- a lexical database for English
- senses in WordNet are generally ordered from most frequent to least frequent based on their counts
- accessed on the Web or downloaded locally

- **Popular sense-tagged corpora:**
  - **SemCor** : https://www.sketchengine.eu/semcorannotated-corpu
  - **Senseval** : https://web.eecs.umich.edu/~mihalcea/senseval/senseval3/tasks.html
  - **Certain SemEval** : http://alt.qcri.org/semeval2015/task13/

WordNet

WordNet 3.0:

- 117,798 nouns, 11,529 verbs, 22,479 adjectives, 4,481 adverbs
- The average noun has 1.23 senses, and the average verb has 2.16 senses

The noun “bass” has 8 senses in WordNet.

1. bass^{1} - (the lowest part of the musical range)
2. bass^{2}, bass part^{1} - (the lowest part in polyphonic music)
3. bass^{3}, basso^{1} - (an adult male singer with the lowest voice)
4. sea bass^{1}, bass^{4} - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass^{1}, bass^{5} - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass^{6}, bass voice^{1}, basso^{2} - (the lowest adult male singing voice)
7. bass^{7} - (the member with the lowest range of a family of musical instruments)
8. bass^{8} - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Figure 19.1 A portion of the WordNet 3.0 entry for the noun *bass*.

Word Sense Disambiguation (WSD)

- **Task :** automatically select the correct sense of a word
  - **Input** : a word in context
  - **Output** : sense of the word

Word Sense Disambiguation (WSD)

- **Task :** automatically select the correct sense of a word
  - **Input** : a word in context
  - **Output** : sense of the word

- **Motivated by many applications**
  - **Machine translation**
    - e.g. translate '*play*' into Persian
    - play the violin = نواختن ویالن
    - play tennis = تنیس بازی کردن
  - **Other uses**
    - Text to speech generation (lead)
    - Accent restoration (cote)
    - Spelling correction (aid/aide)
    - Capitalization restoration (Turkey)

Two variants of WSD task

- **Lexical Sample task**
    - Small pre-selected set of target words (*sentences*, *bank*)
    - And inventory of senses for each word
    - Supervised machine learning: train a classifier for each word
- **All-words task**
    - Every word in an entire text
    - A lexicon with senses for each word
    - Data sparseness: can’t train word-specific classifiers

Two variants of WSD task

All-words task

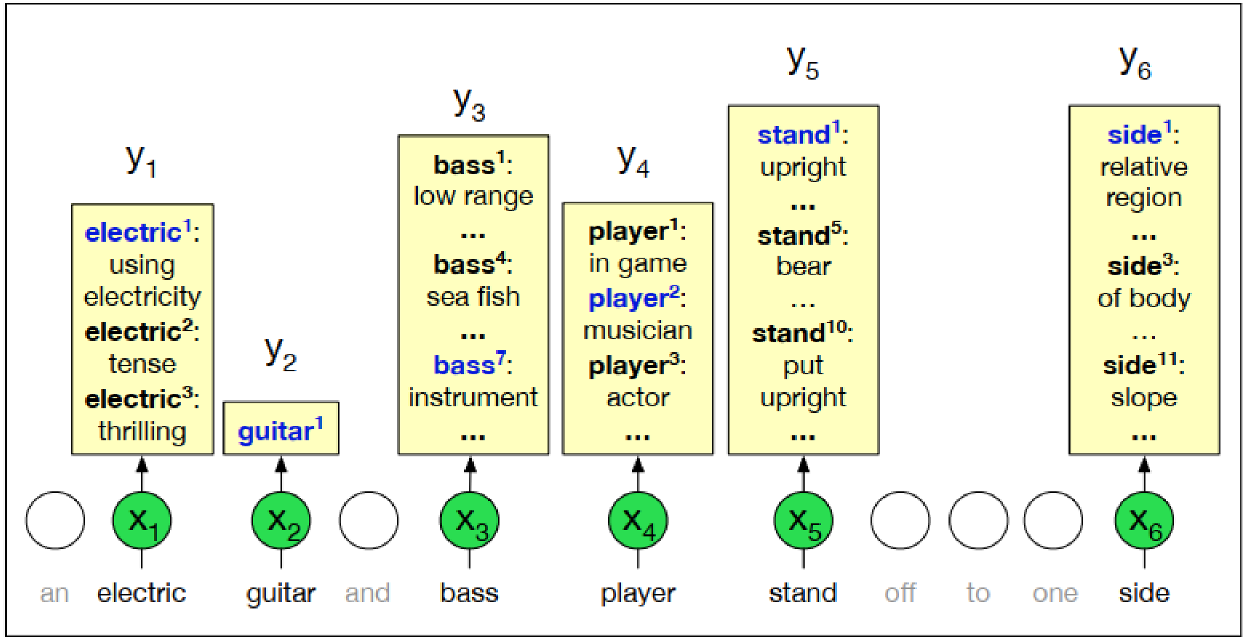


Figure 19.8 The all-words WSD task, mapping from input words (x) to WordNet senses (y). Only nouns, verbs, adjectives, and adverbs are mapped, and note that some words (like *guitar* in the example) only have one sense in WordNet. Figure inspired by Chaplot and Salakhutdinov (2018).

Baseline Performance

Baseline: most frequent sense

- Equivalent to “take first sense” in WordNet
- Does surprisingly well!

Freq	Synset	Gloss
338	$plant^1$, works, industrial plant	buildings for carrying on industrial labor
207	$plant^2$, flora, plant life	a living organism lacking the power of locomotion
2	$plant^3$	something planted secretly for discovery by another
0	$plant^4$	an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience

62% accuracy in this case!

Baseline Performance

Baseline: most frequent sense

- Equivalent to “take first sense” in WordNet
- Does surprisingly well!

Freq	Synset	Gloss
338	$plant^1$, works, industrial plant	buildings for carrying on industrial labor
207	$plant^2$, flora, plant life	a living organism lacking the power of locomotion
2	$plant^3$	something planted secretly for discovery by another
0	$plant^4$	an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience

62% accuracy in this case!

- **Generating sense labeled corpora is quite difficult and expensive**

Dictionary Methods

Baseline: Lesk Algorithm

- Classic
- Powerful
- knowledge-based approach
- Match sentences to dictionary definitions

Intuition: Given the glosses for all possible senses of a word, the gloss that shares the most words with the immediate context of the target word corresponds to the correct sense

Dictionary Methods

Baseline: Simplified Lesk algorithm (Kilgarriff and Rosenzweig, 2000)

```
function SIMPLIFIED LESK(word, sentence) returns best sense of word

  best-sense ← most frequent sense for word
  max-overlap ← 0
  context ← set of words in sentence
  for each sense in senses of word do
    signature ← set of words in the gloss and examples of sense
    overlap ← COMPUTEOVERLAP(signature, context)
    if overlap > max-overlap then
        max-overlap ← overlap
        best-sense ← sense
  end
  return(best-sense)
```

Figure 19.10 The Simplified Lesk algorithm. The COMPUTEOVERLAP function returns the number of words in common between two sets, ignoring function words or other words on a stop list. The original Lesk algorithm defines the *context* in a more complex way.

Dictionary Methods

Baseline: Simplified Lesk algorithm (Kilgarriff and Rosenzweig, 2000)

The **bank** can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

Bank(1)	Gloss	A financial institution that accepts deposits and channels the money into lending activities
	Examples	"he cashed a check at the bank," "that bank holds the mortgage on my home"
Bank(2)	Gloss	Sloping land (especially the slope beside a body of water)
	Examples	"they pulled the canoe up on the bank," "he sat on the bank of the river and watched the currents

Dictionary Methods

Baseline: Simplified Lesk algorithm (Kilgarriff and Rosenzweig, 2000)

The **bank** can guarantee **deposits** will eventually cover future tuition costs because it invests in adjustable-rate **mortgage** securities.

Bank(1)	Gloss	A financial institution that accepts **deposits** and channels the money into lending activities
	Examples	"he cashed a check at the bank," "that bank holds the **mortgage** on my home"
Bank(2)	Gloss	Sloping land (especially the slope beside a body of water)
	Examples	"they pulled the canoe up on the bank," "he sat on the bank of the river and watched the currents

Supervised Machine Learning

- **Supervised machine learning approach:**
    - A training corpus of words tagged in context with their senses
    - Used to train a classifier that can tag words in new texts

Supervised Machine Learning

- **Supervised machine learning approach:**
  - A training corpus of words tagged in context with their senses
  - Used to train a classifier that can tag words in new texts
- **Summary of what we need:**
  - The tag set (“sense inventory”)
  - The training corpus
  - A set of features extracted from the training corpus
  - A classifier

Supervised Machine Learning

- **Supervised machine learning approach:**
  - A training corpus of words tagged in context with their senses
  - Used to train a classifier that can tag words in new texts
- **Summary of what we need:**
  - The tag set ("sense inventory") **-> Dictionary, WordNet**
  - The training corpus
  - A set of features extracted from the training corpus
  - A classifier

The noun "bass" has 8 senses in WordNet.

1. bass^{1} - (the lowest part of the musical range)
2. bass^{2}, bass part^{1} - (the lowest part in polyphonic music)
3. bass^{3}, basso^{1} - (an adult male singer with the lowest voice)
4. sea bass^{1}, bass^{4} - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass^{1}, bass^{5} - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass^{6}, bass voice^{1}, basso^{2} - (the lowest adult male singing voice)
7. bass^{7} - (the member with the lowest range of a family of musical instruments)
8. bass^{8} - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Figure 19.1 A portion of the WordNet 3.0 entry for the noun *bass*.

Supervised Machine Learning

- **Supervised machine learning approach:**
  - A training corpus of words tagged in context with their senses
  - Used to train a classifier that can tag words in new texts
- **Summary of what we need:**
  - The tag set ("sense inventory") **-> Dictionary, WordNet**
  - The training corpus **-> SemCor**
  - A set of features extracted from the training corpus
  - A classifier

Supervised Machine Learning

- **Supervised machine learning approach:**
  - A training corpus of words tagged in context with their senses
  - Used to train a classifier that can tag words in new texts
- **Summary of what we need:**
  - The tag set ("sense inventory") **-> Dictionary, WordNet**
  - The training corpus **-> SemCor**
  - A set of features extracted from the training corpus **-> Feature vectors**
  - A classifier

Feature vectors

- **A simple representation for each observation:**

  (each instance of a target words)

  - Vectors of sets of feature/value pairs
  - Represented as a ordered list of values
  - These vectors represent, e.g., the window of words around the target
  - A training corpus of words tagged in context with their senses

Two kinds of features in the vectors

- **Collocational**
    - Features about words at specific positions near target word
        - Often limited to just word identity and POS

- **Bag-of-words**
    - Features about words that occur anywhere in the window (regardless of position)
        - Typically limited to frequency counts

Feature vectors

- **Example:**

  An electric guitar and **bass** player stand off to one side.

Feature vectors

- **Example:**

An electric guitar and **bass** player stand off to one side.

  - Assume a window of +/-2 from the target

Feature vectors

- **Example:**

  An electric guitar and **bass** player stand off to one side.

  - Assume a window of +/-2 from the target
  - **Collocational features**
    - Position specific information about the words and collocations in window

$$[w_{i-2}, \mathrm{POS}_{i-2}, w_{i-1}, \mathrm{POS}_{i-1}, w_{i+1}, \mathrm{POS}_{i+1}, w_{i+2}, \mathrm{POS}_{i+2}, w_{i-2}^{i-1}, w_{i}^{i+1}]$$

```
[guitar, NN, and, CC, player, NN, stand, VB, and guitar, player stand]
```

    - word 1,2,3 grams in window of ±3 is common

Feature vectors

- **Example:**

  An electric guitar and **bass** player stand off to one side.

  - Assume a window of +/-2 from the target
  - **Bag-of-words features**
    - "an unordered set of words"– position ignored
    - Counts of words occur within the windows
    - First choose a vocabulary
    - Then count how often each of those terms occurs in a given window
    - sometimes just a binary "indicator" 1 or 0

Feature vectors

- **Example:**

  An electric guitar and **bass** player stand off to one side.

  - **Bag-of-words features**
    - **Co-Occurrence:**
      - Assume we have settled on a possible vocabulary of 12 words in "bass" sentences:

        [*fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band*]

      - The vector for "guitar and bass player stand" [0,0,0,1,0,0,0,0,0,0,1,0]

Decision list (Yarowsky)

Collocations	
Correct	Incorrect
• High temperature	• Tall temperature
• Have an experience	• Make an experience
• Heavy rain	• Thick rain

- **Method introduced by Yarowsky (1994)**
- **Two sense per word**
- **Ordered rules: collocation --> sense**
    - Collocations are two or more words that tend to appear frequently together
    - part-of-speech tags (for a window of 3 words on each side, stopping at sentence boundaries)
    - collocation features of words or n-grams of lengths 1, 2, 3 (particular 3)

Decision list (Yarowsky)

Collocations	
Correct	**Incorrect**
• High temperature • Have an experience • Heavy rain	• Tall temperature • Make an experience • Thick rain

- **Method introduced by Yarowsky (1994)**
- **Two sense per word**
- **Ordered rules: collocation --> sense**
    - Collocations are two or more words that tend to appear frequently together
    - part-of-speech tags (for a window of 3 words on each side, stopping at sentence boundaries)
    - collocation features of words or n-grams of lengths 1, 2, 3 (particular 3)
- **Formula**

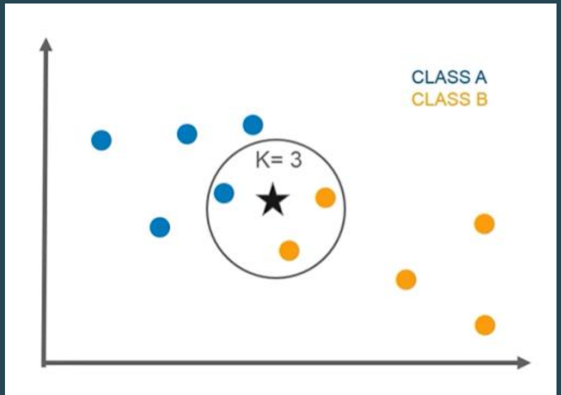
$$\log\left(\frac{p(sense_A|collocation_i)}{p(sense_B|collocation_i)}\right)$$

Classification Features

- **Adjacent words (collocations)**
  - e.g. , bar exam, chocolate bar, bar fight
- **Adjacent parts of speech**
- **Nearby words**
  - e.g. within 10 words
- **Syntactic information**
  - e.g. object of the verb ‘play’
- **Topic of text**
  - One sense per discourse (Gale et al.1992)

Classification Methods

- K-nearest neighbor (memory-based)
- Using Euclidean distance
- Find the k most similar examples and return the majority class for them



Semi-Supervised Learning

Problem: supervised and dictionary-based approaches require large hand-built resources

What if you don't have so much training data?

Semi-Supervised Learning

Problem: supervised and dictionary-based approaches require large hand-built resources

What if you don't have so much training data?

Solution: Bootstrapping

- Generalize from a very small hand-labeled seed-set

Bootstrapping

- **Semi-supervised**
- **One of the most influential algorithms**

How it works?

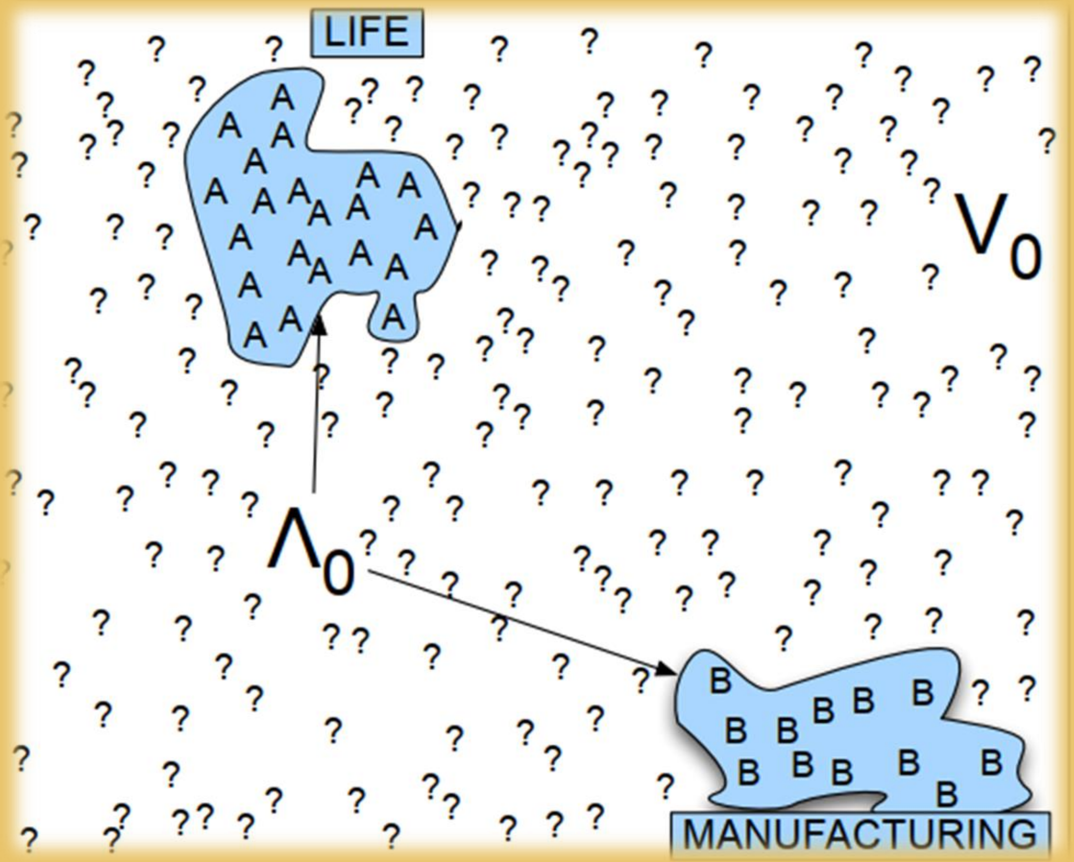
- Start with two sense and seeds for each sense
    - e.g., plant1: leaf, plant2: factory
- Use these seeds to label the data using a supervised classifier (decision list)
- Add some of the newly labeled examples to the training data

Bootstrapping

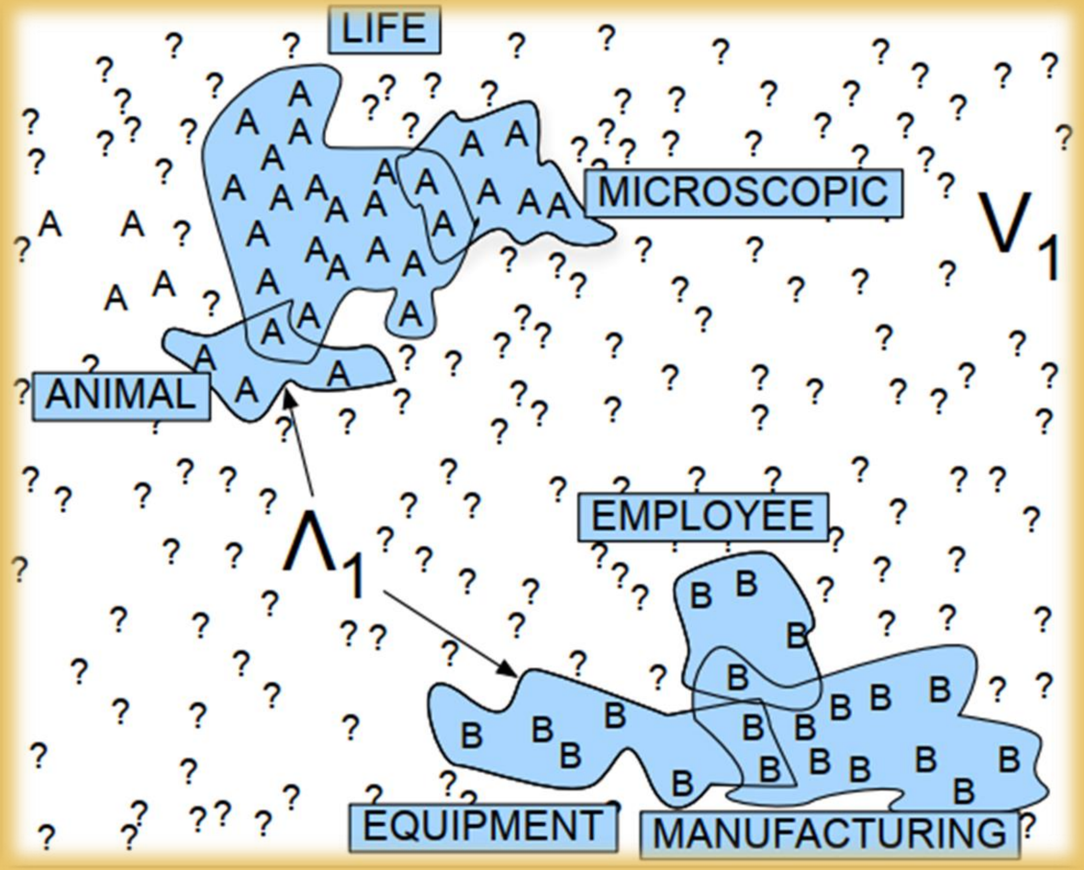
Summary: generating seeds

1. **Hand labeling**
2. **One sense per collocation**
   - A word reoccurring in collocation with the same word will almost surely have the same sense.
   - The first one
3. **One sense per discourse**
   - The sense of a word is highly consistent within a document Yarowsky(1994)
   - (At least for non-function words, and especially topic--specific words)

Stages in the Yarowsky bootstrapping algorithm for word ‘plant’



Stages in the Yarowsky bootstrapping algorithm for word 'plant'



Training data for WSD

Senseval/Semcor:

- http://www.senseval.org/senseval3
- Lexical sample
- All words
- Available for many language

Senseval-1 Evaluation

Metrics:

- A = number of assigned senses
- C = number of words assigned correct senses
- T = total number of test words
- Precision = C/A; Recall = C/T

Result:

- Best recall around 77P/77R
- Human lexicographer 97P/96R
- Most common sense 57P/50R

Thanks ☺