Word Senses and WordNet

(Word Sense Disambiguation)

Midterm

Minimum Value	54.00
Maximum Value	100.00
Average	83.325
Median	84.50

90 - 100	11
80 - 89	19
70 - 79	5
60 - 69	4
50 - 59	1

Review

- Lexical Semantics / Relationships
- WordNet
- Word Similarity via WordNet Thesaurus

Word Sense Disambiguation (WSD)

- Given
  - A word in context,
  - A fixed inventory of potential word senses
- Decide which sense of the word this is
- What set of senses?
  - English-to-Spanish MT: set of Spanish translations
  - Speech Synthesis: homographs like *bass* and *bow*
  - In general: the senses in a thesaurus like WordNet

TheWordNet entry for the noun ***bat*** has the following distinct senses.

Cluster these senses by using the definitions of homonymy and polysemy.

- bat#1: nocturnal mouselike mammal
- bat#2: (baseball) a turn trying to get a hit
- bat#3: a small racket. . . for playing squash
- bat#4: the club used in playing cricket
- bat#5: a club used for hitting a ball in various games

Two Variants of WSD

- Lexical Sample task
  - Small pre-selected set of target words
  - And inventory of senses for each word
  - Typically supervised ML: classifier per 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
  - ~Like part-of-speech tagging
    - Except each lemma has its own tagset
    - Less human agreement so upper bound is lower

Word-in-Context (Simpler) Task

F	There's a lot of trash on the **bed** of the river — I keep a glass of water next to my **bed** when I sleep
F	**Justify** the margins — The end **justifies** the means
T	**Air** pollution — Open a window and let in some **air**
T	The expanded **window** will give us time to catch the thieves — You have a two-hour **window** of clear weather to finish working on the lawn

Figure 19.11 Positive (T) and negative (F) pairs from the WiC dataset (Pilehvar and Camacho-Collados, 2019).

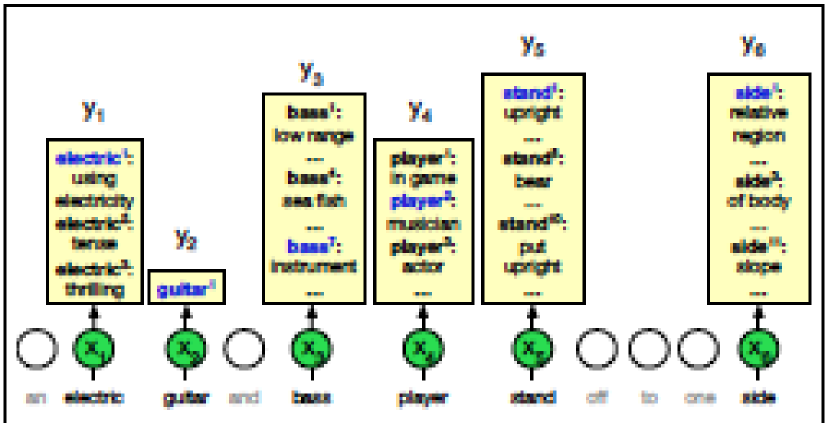


Figure 19.8 The all-words WSD task, mapping from input words (x) to WordNet senses (y). Only nouns, verbs, adjectives, and adverbs are mapped, and note that some words (like *guitar* in the example) only have one sense in WordNet. Figure inspired by Chaplot and Salakhutdinov (2018).

Approaches

- Supervised machine learning
- "Unsupervised"
  - Thesaurus/Dictionary-based techniques
  - (Selectional Association, next chapter)
- Lightly supervised
  - Bootstrapping

Supervised Machine Learning Approaches

- Supervised machine learning approach
  - Training corpus depends on task
  - Train a classifier that can tag words in new text
  - Just as we saw for part-of-speech tagging
- What do we need?
  - Tag set ("sense inventory")
  - Training corpus (words tagged in context with sense)
  - Embeddings or set of features extracted from the training corpus
  - A classifier

Supervised WSD: WSD Tags

- What's a tag?

  A dictionary sense?
- For example, for WordNet an instance of "bass" in a text has 8 possible tags or labels (bass1 through bass8).

Bass in WordNet

- The noun bass has 8 senses in WordNet
  - bass - (the lowest part of the musical range)
  - bass, bass part - (the lowest part in polyphonic music)
  - bass, basso - (an adult male singer with the lowest voice)
  - sea bass, bass - (flesh of lean-fleshed saltwater fish of the family Serranidae)
  - freshwater bass, bass - (any of various North American lean-fleshed freshwater fishes especially of the genus Micropterus)
  - bass, bass voice, basso - (the lowest adult male singing voice)
  - bass - (the member with the lowest range of a family of musical instruments)
  - bass -(nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Sense Tags for Bass

WordNet Sense	Spanish Translation	Roget Category	Target Word in Context
$bass^4$	lubina	FISH/INSECT	. . . fish as Pacific salmon and striped **bass** and. . .
$bass^4$	lubina	FISH/INSECT	. . . produce filets of smoked **bass** or sturgeon. . .
$bass^7$	bajo	MUSIC	. . . exciting jazz **bass** player since Ray Brown. . .
$bass^7$	bajo	MUSIC	. . . play **bass** because he doesn't have to solo. . .

What kind of Corpora?

- Lexical sample task:
  - *Line-hard-serve* corpus - 4000 examples of each
  - *Interest* corpus - 2369 sense-tagged examples
- All words:
  - **Semantic concordance**: a corpus in which each open-class word is labeled with a sense from a specific dictionary/thesaurus.
    - SemCor: 234,000 words from Brown Corpus, manually tagged with WordNet senses
    - SENSEVAL-3 competition corpora - 2081 tagged word tokens

SemCor

<wf pos=PRP>**He**</wf>
<wf pos=VB lemma=recognize wnsn=4 lexsn=2:31:00::>**recognized**</wf>
<wf pos=DT>**the**</wf>
<wf pos=NN lemma=gesture wnsn=1 lexsn=1:04:00::>**gesture**</wf>
<punc>.</punc>

16

(Contextual) Embedding Algorithms

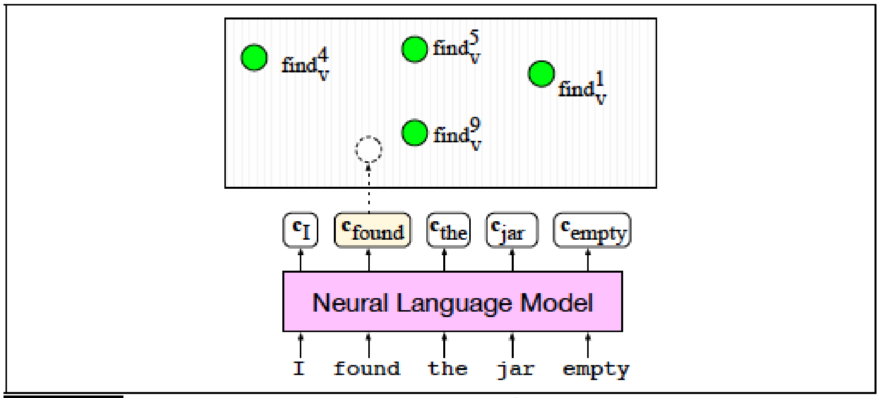


Figure 19.9 The nearest-neighbor algorithm for WSD. In green are the contextual embeddings precomputed for each sense of each word; here we just show a few of the senses for *find*. A contextual embedding is computed for the target word *found*, and the and then the nearest neighbor sense (in this case $\mathbf{find}_n^9$) would be chosen. Figure inspired by Loureiro and Jorge (2019).

Contextual Embedding Algorithms (continued)

Training

- Embed each token in a sense-labeled training corpus
- Average each token of each sense of each word to produce a sense embedding
  - – See arxiv article linked in syllabus for Contextual extension of embeddings

Testing

- Compare test embedding with training embeddings
- Return sense of nearest neighbor based on a similarity metric such as cosine

What about unseen test words?

- Backoff via most frequent sense (first sense in WordNet)
- Imputation via WordNet taxonomy and supersenses

Frequency-based WSD

- WordNet first sense heuristic, about 60-70% accuracy
- To improve, need context
  - Selectional restrictions
  - “Topic”

Imputation

More formally, for each missing sense in WordNet $\hat{s} \in W$, let the sense embeddings for the other members of its synset be $S_{\hat{s}}$, the hypernym-specific synset embeddings be $H_{\hat{s}}$, and the lexicographic (supersense-specific) synset embeddings be $L_{\hat{s}}$. We can then compute the sense embedding for $\hat{s}$ as follows:

$$\text{if} \quad |S_{\hat{s}}| > 0, \quad \mathbf{v}_{\hat{s}} = \frac{1}{|S_{\hat{s}}|} \sum \mathbf{v}_s, \forall \mathbf{v}_s \in S_{\hat{s}} \tag{19.14}$$

$$\text{else if} \quad |H_{\hat{s}}| > 0, \quad \mathbf{v}_{\hat{s}} = \frac{1}{|H_{\hat{s}}|} \sum \mathbf{v}_{syn}, \forall \mathbf{v}_{syn} \in H_{\hat{s}} \tag{19.15}$$

$$\text{else if} \quad |L_{\hat{s}}| > 0, \quad \mathbf{v}_{\hat{s}} = \frac{1}{|L_{\hat{s}}|} \sum \mathbf{v}_{syn}, \forall \mathbf{v}_{syn} \in L_{\hat{s}} \tag{19.16}$$

Review

- Project comments
  - Code formats
  - Paper formats
  - Baselines, measuring progress (can't compare to published paper)
  - Provided data (no negative sampling needed)
  - GPU
- ArXiv and NYTimes links
- Last class
  - Tasks
  - Embedding approach to WSD

Combining Embeddings/Thesaurus

- Addressing problem of antonyms
- E.g., counterfitting
  - After embeddings are trained, use a thesaurus to learn a second mapping that shifts antonyms apart and synonyms closer

	Before counterfitting			After counterfitting		
east	west	north	south	eastward	eastern	easterly
expensive	pricey	cheaper	costly	costly	pricy	overpriced
British	American	Australian	Britain	Brits	London	BBC

Figure 19.12 The nearest neighbors in GloVe to *east*, *expensive*, and *British* include antonyms like *west*. The right side showing the improvement in GloVe nearest neighbors after the counterfitting method (Mrkšić et al., 2016).

Non-Neural: What Kind of Features?

- Weaver (1955) "If one examines the words in a book, one at a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words. […] But if one lengthens the slit in the opaque mask, until one can see not only the central word in question but also say N words on either side, then if N is large enough one can unambiguously decide the meaning of the central word. […] The practical question is : `What minimum value of N will, at least in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?'"

•S: (n) **dish** (a piece of dishware normally used as a container for holding or serving food) *"we gave them a set of dishes for a wedding present"*

•S: (n) **dish** (a particular item of prepared food) *"she prepared a special dish for dinner"*

….

dishes

- washing *dishes* he
- simple *dishes* including
- convenient *dishes* to
- of *dishes* and

bass

- free *bass* with
- pound *bass* of
- and *bass* player
- his *bass* while

dishes

- includes washing *dishes* he says
- several simple *dishes* including braised
- and convenient *dishes* to fix
- variety of *dishes* and regional

bass

- the free *bass* with ease
- 52 pound *bass* of a
- guitar and *bass* player stand
- caught his *bass* while fishing

- "In our house, everybody has a career and none of them includes washing *dishes*," he says.
- In her tiny kitchen at home, Ms. Chen works efficiently, stir-frying several simple *dishes*, including braised pig's ears and chcken livers with green peppers.
- Post quick and convenient *dishes* to fix when you're in a hurry.
- Japanese cuisine offers a great variety of *dishes* and regional specialties

- We need more good teachers – right now, there are only a half a dozen who can play the free *bass* with ease.
- Though still a far cry from the lake's record 52 pound *bass* of a decade ago, "you could fillet these fish again, and that made people very, very happy." Mr. Paulson says.
- An electric guitar and *bass* player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations again.
- Lowe caught his *bass* while fishing with pro Bill Lee of Killeen, Texas, who is currently in 144th place with two bass weighing 2-09.

Feature Vectors

- A simple representation for each observation (each instance of a target word)
  - Vectors of sets of feature/value pairs
    - I.e. files of comma-separated values
  - These vectors should represent the window of words around the target

  How big should that window be?

What sort of Features?

- **Collocational** features and **bag-of-words** features
  - **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

Example

- Example text (WSJ)
  - An electric guitar and **bass** player stand off to one side not really part of the scene, just as a sort of nod to gringo expectations perhaps
  - Assume a window of +/- 2 from the target

Collocations

- Position-specific information about the words in the window
- guitar and bass player stand
  - [guitar, NN, and, CC, player, NN, stand, VB]
  - $Word_{n-2,}$ $POS_{n-2,}$ $word_{n-1,}$ $POS_{n-1,}$ $Word_{n+1}$ POS_{n+1}…
  - In other words, a vector consisting of
  - [position n word, position n part-of-speech…]

Bag of Words

- Information about what words occur within the window
- First derive a set of terms to place in the vector
- Then note how often each of those terms occurs in a given window

Co-Occurrence Example

- Assume we've 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

Co-Occurrence Example

- Assume we've 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]

Classifiers

- Once we cast the WSD problem as a classification problem, many techniques possible
  - Naïve Bayes
  - Decision lists
  - Decision trees
  - Neural nets
  - Support vector machines
  - Nearest neighbor methods…

Classifiers

- Choice of technique, in part, depends on the set of features that have been used
  - Some techniques work better/worse with features with numerical values
  - Some techniques work better/worse with features that have large numbers of possible values
    - For example, the feature **the word to the left** has a fairly large number of possible values

Classification Methods: Supervised Machine Learning

- *Input:*
  - a word *w in a text window d (which we'll call a "document")*
  - a fixed set of classes $C = \{c_1, c_2, \ldots, c_J\}$
  - A training set of *m* hand-labeled text windows again called "documents" $(d_1,c_1),\ldots,(d_m,c_m)$
- *Output:*
  - a learned classifier $\gamma: d \rightarrow c$

39

Naïve Bayes

- $\hat{s} = \arg\max_{s \in S} p(s|V)$, or $\arg\max_{s \in S} \frac{p(V|s)\,p(s)}{p(V)}$
- Where s is one of the senses S possible for a word w and V the input vector of feature values for w
- Assume features ***independent***, so probability of V is the product of probabilities of each feature, given s, so $p(V|s) = \prod_{j=1}^{n} p(v_j|s)$
- p(V) same for any ŝ
- Then $\hat{s} = \arg\max_{s \in S} p(s) \prod_{j=1}^{n} p(v_j|s)$

- How do we estimate $p(s)$ and $p(v_j|s)$?

- How do we estimate $p(s)$ and $p(v_j|s)$?
  - $p(s_i)$ is max. likelihood estimate from a sense-tagged corpus ($count(s_i,w_j)/count(w_j)$) – how likely is bank to mean 'financial institution' over all instances of bank?
  - $P(v_j|s)$ is max. likelihood of each feature given a candidate sense ($count(v_j,s)/count(s)$) – how likely is the previous word to be 'river' when the sense of bank is 'financial institution'
- Calculate $\hat{s} = \arg\max_{s \in S} p(s) \prod_{j=1}^{n} p(v_j|s)$ for each possible sense and take the highest scoring sense as the most likely choice

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w \mid c) = \frac{count(w,c)+1}{count(c)+|V|}$$

	Doc	Words	Class
Training	1	fish smoked fish	f
	2	fish line	f
	3	fish haul smoked	f
	4	guitar jazz line	g
Test	5	line guitar jazz jazz	?

V = {fish, smoked, line, haul, guitar, jazz}

Priors:

$P(f) = \frac{3}{4}$

$P(g) = \frac{1}{4}$

Conditional Probabilities:

P(line|*f*) = (1+1) / (8+6) = 2/14

P(guitar|*f*) = (0+1) / (8+6) = 1/14

P(jazz|*f*) = (0+1) / (8+6) = 1/14

P(line|*g*) = (1+1) / (3+6) = 2/9

P(guitar|*g*) = (1+1) / (3+6) = 2/9

P(jazz|*g*) = (1+1) / (3+6) = 2/9

Choosing a class:

$P(f|d5) \propto 3/4 * 2/14 * (1/14)^2 * 1/14 \approx 0.00003$

$P(g|d5) \propto 1/4 * 2/9 * (2/9)^2 * 2/9 \approx 0.0006$

43

WSD Evaluations and Baselines

- Intrinsic versus Extrinsic evaluation
  - Exact match **accuracy**
    - % of words tagged identically with manual sense tags
- Baselines: most frequent sense, one sense per discourse, Lesk algorithm

Naïve Bayes Evaluation

- On a corpus of examples of uses of the word line, naïve Bayes achieved about 73% correct
- Is this good?

Most Frequent Sense

- Wordnet senses are ordered in frequency order
- So "most frequent sense" in WordNet = "take the first sense"
- Sense frequencies come from SemCor

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

Ceiling

- Human inter-annotator agreement
  - Compare annotations of two humans
  - On same data
  - Given same tagging guidelines
- Human agreements on all-words corpora with WordNet style senses
  - 75%-80%

Way to easily get training data?

- Wikipedia
  - Link ambigious words to different articles
  - Then map articles to sense inventory
  - BabelNet
    - live.babelnet.org

In 1834, Sumner was admitted to the [[**bar (law)**|**bar**]] at the age of twenty-three, and entered private practice in Boston.

It is danced in 3/4 time (like most waltzes), with the couple turning approx. 180 degrees every [[**bar (music)**|**bar**]].

Jenga is a popular beer in the [[**bar (establishment)**|**bar**]]s of Thailand.

Unsupervised Methods: Dictionary/Thesaurus Methods

- The Lesk Algorithm

Simplified Lesk

- Match dictionary entry of sense that best matches context

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

given the following two WordNet senses:

$bank^1$	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"

Simplified Lesk

- Match dictionary entry of sense that best matches context: bank1 (deposits, mortgage)

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

given the following two WordNet senses:

$bank^1$	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"

Original Lesk: pine cone

- Compare entries for each context word for overlap

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

Original Lesk: pine cone

- Compare entries for each context word for overlap
- Cone3 selected: evergreen, tree

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

"Time flies like an arrow" – what are the correct senses?

- time#n#5 (the continuum of experience in which events pass from the future through the present to the past)
- time#v#1 (measure the time or duration of an event or action or the person who performs an action in a certain period of time) "he clocked the runners"

- flies#n#1 (two-winged insects characterized by active flight)
- flies#v#8 (pass away rapidly) "Time flies like an arrow"; "Time fleeing beneath him"

- like#v#4 (feel about or towards; consider, evaluate, or regard) "How did you like the President's speech last night?"
- like#a#1 (resembling or similar; having the same or some of the same characteristics; often used in combination) "suits of like design"; "a limited circle of likeminds"; "members of the cat family have like dispositions"; "as like as two peas in a pod"; "doglike devotion"; "a dreamlike quality"

Try the original algorithm on *"Time flies like an arrow"* using WordNet senses below. Assume that the words are to be disambiguated one at a time, from left to right,
and that the results from earlier decisions are used later in the process.

- time#n#5 (the continuum of experience in which events pass from the future through the present to the past)
- time#v#1 (measure the time or duration of an event or action or the person who performs an action in a certain period of time) "he clocked the runners"

- flies#n#1 (two-winged insects characterized by active flight)
- flies#v#8 (pass away rapidly) "Time flies like an arrow"; "Time fleeing beneath him"

- like#v#4 (feel about or towards; consider, evaluate, or regard) "How did you like the President's speech last night?"
- like#a#1 (resembling or similar; having the same or some of the same characteristics; often used in combination) "suits of like design"; "a limited circle of likeminds"; "members of the cat family have like dispositions"; "as like as two peas in a pod"; "doglike devotion"; "a dreamlike quality"

"Time flies like an arrow"

- time#n#5 (the continuum of experience in which events **pass** from the future through the present to the past)
- time#v#1 (measure the **time** or duration of an event or action or the person who performs an action in a certain period of time) "he clocked the runners"

- flies#n#1 (two-winged insects characterized by active flight)
- flies#v#8 (**pass** away rapidly) "**Time** flies like an arrow"; "**Time** fleeing beneath him"

- like#v#4 (feel about or towards; consider, evaluate, or regard) "How did you like the President's speech last night?"
- like#a#1 (resembling or similar; having the same or some of the same characteristics; often used in combination) "suits of like design"; "a limited circle of likeminds"; "members of the cat family have like dispositions"; "as like as two peas in a pod"; "doglike devotion"; "a dreamlike quality"

- **Time WSD : tie, backoff to most frequent, but can't because POS differ**

"Time flies like an arrow"

- time#n#5 (the continuum of experience in which events **pass** from the future through the present to the past)
- time#v#1 (measure the **time** or duration of an event or action or the person who performs an action in a certain period of time) "he clocked the runners"

- flies#n#1 (**two**-winged insects characterized by active flight)
- flies#v#8 (**pass** away rapidly) "**Time** flies **like** an arrow"; "**Time** fleeing beneath him"

- like#v#4 (feel about or towards; consider, evaluate, or regard) "How did you **like** the President's speech last night?"
- like#a#1 (resembling or similar; having the same or some of the same characteristics; often used in combination) "suits of like design"; "a limited circle of likeminds"; "members of the cat family have **like** dispositions"; "as **like** as **two** peas in a pod"; "doglike devotion"; "a dreamlike quality"

- **Flies WSD: select verb**

Corpus Lesk

- Add corpus examples to glosses and examples
- The best performing variant

The Corpus Lesk algorithm

- Assumes we have some sense-labeled data (like SemCor)
- Take all the sentences with the relevant word sense:

  These short, "streamlined" meetings usually are sponsored by local ***banks[1]****, Chambers of Commerce, trade associations, or other civic organizations.*
- Now add these to the gloss + examples for each sense, call it the “signature” of a sense.
- Choose sense with most word overlap between context and signature.

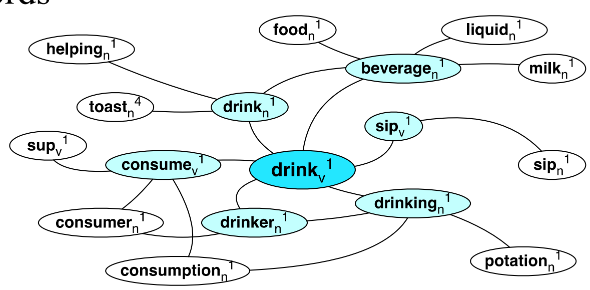
Corpus Lesk: IDF weighting

- Instead of just removing function words
  - Weigh each word by its `promiscuity’ across documents
  - Down-weights words that occur in every `document’ (gloss, example, etc)
  - These are generally function words, but is a more fine-grained measure
- Weigh each overlapping word by **inverse document frequency**

60

Graph-based methods

- First, WordNet can be viewed as a graph
  - senses are nodes
  - relations (hypernymy, meronymy) are edges
  - Also add edge between word and unambiguous gloss words



61

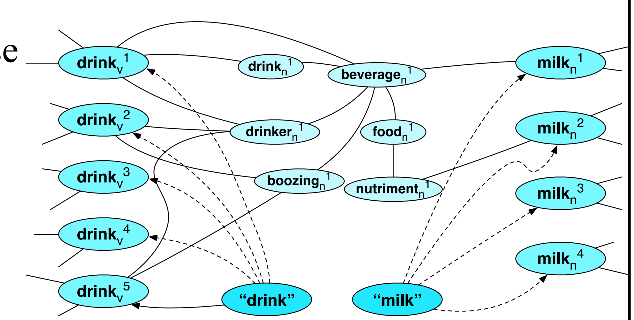
How to use the graph for WSD

- Insert target word and words in its sentential context into the graph, with directed edges to their senses

“She drank some milk”

- Now choose the *most central* sense

Add some probability to “drink” and “milk” and compute node with highest “pagerank”



62

Semi-Supervised Bootstrapping

- What if you don't have enough data or hand-built resources to train a system…
- Bootstrap
  - Pick a word that you as an analyst think will co-occur with your target word in particular sense
  - *Grep* through your corpus for your target word and the hypothesized word
  - Assume that the target tag is the right one
  - Generalize from a small hand-labeled seed set

Bootstrapping

- For bass
  - Assume play occurs with the music sense and fish occurs with the fish sense

Sentences Extracts for bass and player

We need more good teachers – right now, there are only a half a dozen who can **play** the free **bass** with ease.

An electric guitar and **bass player** stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.

When the New Jersey Jazz Society, in a fund-raiser for the American Jazz Hall of Fame, honors this historic night next Saturday, Harry Goodman, Mr. Goodman's brother and **bass player** at the original concert, will be in the audience with other family members.

The researchers said the worms spend part of their life cycle in such **fish** as Pacific salmon and striped **bass** and Pacific rockfish or snapper.

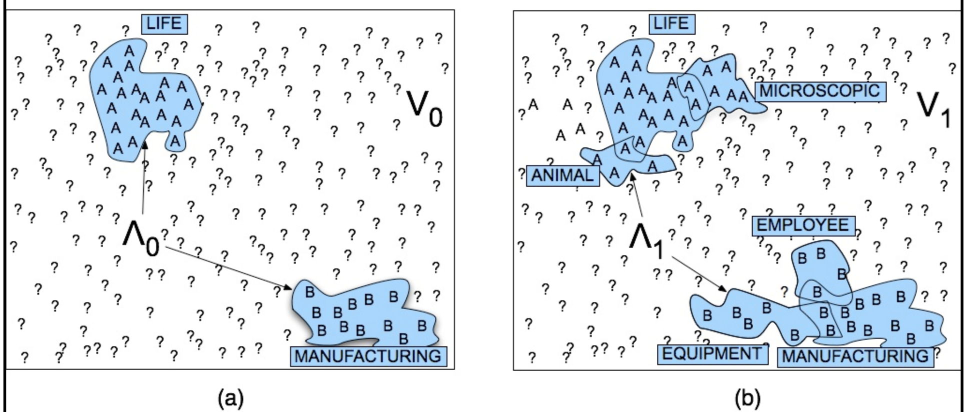
And it all started when **fish**ermen decided the striped **bass** in Lake Mead were too skinny.

Though still a far cry from the lake's record 52-pound **bass** of a decade ago, "you could fillet these **fish** again, and that made people very, very happy," Mr. Paulson says.

Where do the seeds come from?

1) Hand labeling
2) "One sense per discourse":
   - The sense of a word is highly consistent within a document - Yarowsky (1995)
   - True for topic-dependent words
   - Not so true for other POS like adjectives and verbs, e.g. make, take
   - Krovetz (1998) "More than one sense per discourse" not true at all once you move to fine-grained senses
3) One sense per collocation:
   - A word recurring in collocation with the same word will almost surely have the same sense

Stages in Yarowsky Bootstrapping Algorithm



Issues

- Given these general ML approaches, how many classifiers do I need to perform WSD robustly
  - One for each ambiguous word in the language
- How do you decide what set of tags/labels/senses to use for a given word?
  - Depends on the application

WordNet ‘bass’

- Tagging with this set of senses is an impossibly hard task that’s probably overkill for any realistic application

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

Word Sense Induction

- Training

1. For each token w_i of word w in a corpus, compute a context vector **c**.
2. Use a **clustering algorithm** to **cluster** these word-token context vectors **c** into a predefined number of groups or clusters. Each cluster defines a sense of w.
3. Compute the **vector centroid** of each cluster. Each vector centroid $\mathbf{s_j}$ is a **sense vector** representing that sense of w.

- Testing on token t of word w

1. Compute a context vector **c** for t.
2. Retrieve all sense vectors s_j for w.
3. Assign t to the sense represented by the sense vector s_j that is closest to t.

History of Senseval

- ACL-SIGLEX workshop (1997)
- SENSEVAL-I (1998)
  - Lexical Sample for English, French, and Italian
- SENSEVAL-II (Toulouse, 2001)
  - Lexical Sample and All Words
- SENSEVAL-III (2004)
- SENSEVAL-IV -> SEMEVAL (2007)
- Newer: SEM, First Joint Conference on Lexical and Computational Semantics

2012

- SEM: 1st Conf. on Lexical & Computational Semantics
- SemEval: International Workshop on Semantic Evaluations
  - 1. English Lexical Simplification
  - 2. Measuring Degrees of Relational Similarity
  - 3. Spatial Role Labeling
  - 4. Evaluating Chinese Word Similarity
  - 5. Chinese Semantic Dependency Parsing
  - 6. Semantic Textual Similarity
  - 7. COPA: Choice Of Plausible Alternatives An evaluation of
  - common-sense causal reasoning
  - 8. Cross-lingual Textual Entailment for Content Synchronization

WSD Performance

- Varies widely depending on how difficult the disambiguation task is
- Accuracies of over 90% are commonly reported on some of the classic, often fairly easy, WSD tasks (pike, star, interest)
- Senseval brought careful evaluation of difficult WSD (many senses, different POS)
- Senseval 1: more fine grained senses, wider range of types:
  - Overall: about 75% accuracy
  - Nouns: about 80% accuracy
  - Verbs: about 70% accuracy

Summary

- Word Sense Disambiguation: choosing correct sense in context
- Applications: MT, QA, etc.
- Three classes of Methods
  - Supervised Machine Learning: Naive Bayes classifier
  - Thesaurus/Dictionary Methods
  - Semi-Supervised Learning
- Main intuition
  - There is lots of information in a word's context
  - Simple algorithms based just on word counts can be surprisingly good

74