CS674 Natural Language Processing

- Last class
  - Intro to lexical semantics

- Today
  - Lexical semantic resources: WordNet
  - Word sense disambiguation
    » Dictionary-based approaches
    » Supervised machine learning methods
    » Issues for WSD evaluation

WordNet

- Handcrafted database of lexical relations
- Three separate databases: nouns; verbs; adjectives and adverbs
- Each database is a set of lexical entries (according to unique orthographic forms)
  - Set of senses associated with each entry

| Category	Unique Forms	Number of Senses
Noun	94474	116317
Verb	10319	22066
Adjective	20170	29881
Adverb	4546	5677

Sample entry

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

Distribution of senses

- Zipf distribution of senses
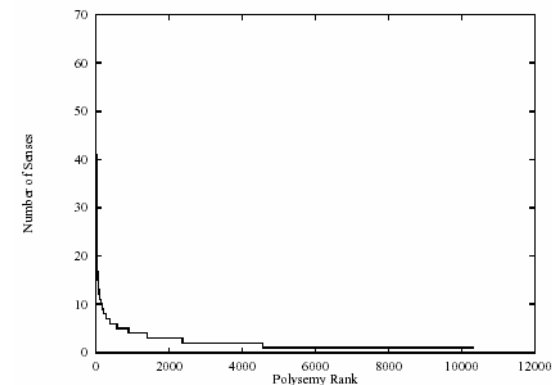
WordNet relations

- Nouns

Relation	Definition	Example
Hypernym	From concepts to superordinates	breakfast → meal
Hyponym	From concepts to subtypes	meal → lunch
Has-Member	From groups to their members	faculty → professor
Member-Of	From members to their groups	copilot → crew
Has-Part	From wholes to parts	table → leg
Part-Of	From parts to wholes	course → meal
Antonym	Opposites	leader → follower

- Verbs

Relation	Definition	Example
Hypernym	From events to superordinate events	fly → travel
Troponym	From events to their subtypes	walk → stroll
Entails	From events to the events they entail	snore → sleep
Antonym	Opposites	increase ⟺ decrease

- Adjectives/adverbs

Relation	Definition	Example
Antonym	Opposite	heavy ⟺ light
Adverb	Opposite	quickly ⟺ slowly

CS674 Natural Language Processing

- Last class
  - Intro to lexical semantics

- Today
  - Lexical semantic resources: WordNet
  - ⟹ Word sense disambiguation
    » Dictionary-based approaches
    » Supervised machine learning methods
    » Issues for WSD evaluation

Word sense disambiguation

- Given a *fixed* set of senses is associated with a lexical item, determine which of them applies to a particular instance of the lexical item
- Two fundamental approaches
  - WSD occurs during semantic analysis as a side-effect of the elimination of ill-formed semantic representations
  - ⟹ Stand-alone approach
    » WSD is performed independent of, and prior to, compositional semantic analysis
    » Makes minimal assumptions about what information will be available from other NLP processes
    » Applicable in large-scale practical applications

Dictionary-based approaches

- Rely on machine readable dictionaries
- Initial implementation of this kind of approach is due to Michael Lesk (1986)
  - Given a word W to be disambiguated
    » Retrieve all of the sense definitions, S, for W from the MRD
    » Compare each s in S to the dictionary definitions of all the remaining words in the context
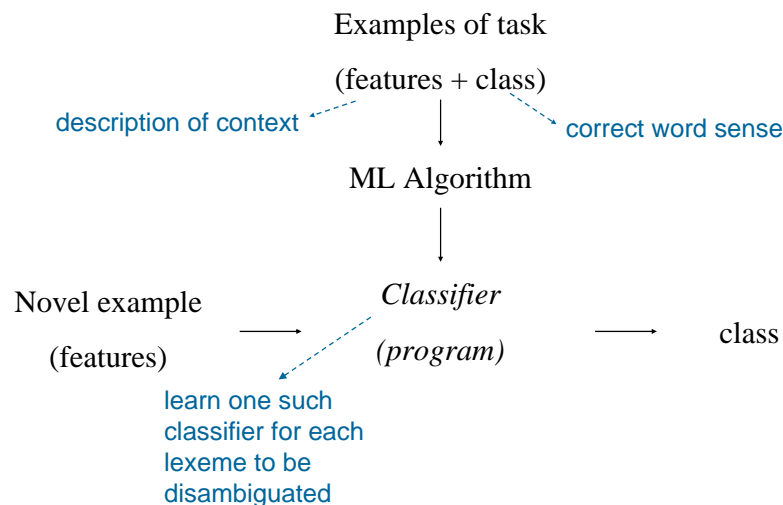   » Select the sense s with the most overlap with (the definitions of) these context words

Example

- Word: *cone*
- Context: *pine cone*
- Sense definitions

  pine 1 kind 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

- Accuracy of 50-70% on short samples of text from *Pride and Prejudice* and an AP newswire article.

Machine learning approaches

- Machine learning methods
  - Supervised inductive learning
  - Bootstrapping
  - Unsupervised
- Emphasis is on acquiring the knowledge needed for the task from data, rather than from human analysts.

Inductive ML framework

Examples of task

(features + class)

description of context   correct word sense

ML Algorithm

Novel example   *Classifier*

(features)   *(program)*   class

learn one such classifier for each lexeme to be disambiguated

Running example

*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.*

1 Fish sense

2 Musical sense

3 …

Feature vector representation

- **target:** the word to be disambiguated
- **context** : portion of the surrounding text
  - Select a "window" size
  - Tagged with part-of-speech information
  - Stemming or morphological processing
  - Possibly some partial parsing
- Convert the context (and target) into a set of features
  - Attribute-value pairs
    » Numeric or nominal values

Collocational features

- Encode information about the lexical inhabitants of *specific* positions located to the left or right of the target word.
  - E.g. the word, its root form, its part-of-speech

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

  - [guitar, NN1, and, CJC, player, NN1, stand, VVB]

Co-occurrence features

- Encodes information about neighboring words, ignoring exact positions.
  - **Features**: the words themselves (or their roots)
  - **Values**: number of times the word occurs in a region surrounding the target word
  - Select a small number of frequently used content words for use as features
    » 12 most frequent content words from a collection of *bass* sentences drawn from the WSJ: *fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band*
    » Co-occurrence vector (window of size 10) for the previous example:
      [0,0,0,1,0,0,0,0,0,0,1,0]

Decision list classifiers

- Decision lists: equivalent to simple case statements.
  - Classifier consists of a sequence of tests to be applied to each input example/vector; returns a word sense.
- Continue only until the first applicable test.
- Default test returns the majority sense.

Decision list example

- Binary decision: fish *bass* vs. musical *bass*

Rule		Sense
fish within window	⇒	**bass**[1]
striped bass	⇒	**bass**[1]
guitar within window	⇒	**bass**[2]
bass player	⇒	**bass**[2]
piano within window	⇒	**bass**[2]
tenor within window	⇒	**bass**[2]
sea bass	⇒	**bass**[1]
play/V bass	⇒	**bass**[2]
river within window	⇒	**bass**[1]
violin within window	⇒	**bass**[2]
salmon within window	⇒	**bass**[1]
on bass	⇒	**bass**[2]
bass are	⇒	**bass**[1]

Learning decision lists

- Consists of *generating* and *ordering* individual tests based on the characteristics of the training data
- Generation: every feature-value pair constitutes a test
- Ordering: based on accuracy on the training set

$$abs\left(\log \frac{P(Sense_1 \mid f_i = v_j)}{P(Sense_2 \mid f_i = v_j)} \right)$$

- Associate the appropriate sense with each test

WSD Evaluation

- Corpora:
  - *line* corpus
  - Yarowsky's 1995 corpus
    » 12 words (plant, space, bass, …)
    » ~4000 instances of each
  - Ng and Lee (1996)
    » 121 nouns, 70 verbs (most frequently occurring/ambiguous); WordNet senses
    » 192,800 occurrences
  - SEMCOR (Landes et al. 1998)
    » Portion of the Brown corpus tagged with WordNet senses
  - SENSEVAL (Kilgarriff and Rosenzweig, 2000)
    » Annual performance evaluation conference
    » Provides an evaluation framework (Kilgarriff and Palmer, 2000)
- Baseline: most frequent sense

WSD Evaluation

- Metrics
  - Precision
    » Nature of the senses used has a huge effect on the results
    » E.g. results using coarse distinctions cannot easily be compared to results based on finer-grained word senses
  - Partial credit
    » Worse to confuse musical sense of *bass* with a fish sense than with another musical sense
    » Exact-sense match → full credit
    » Select the correct broad sense → partial credit
    » Scheme depends on the organization of senses being used

SENSEVAL-2

- Three tasks
  - Lexical sample
  - All-words
  - Translation
- 12 languages
- Lexicon
  - SENSEVAL-1: from HECTOR corpus
  - SENSEVAL-2: from WordNet 1.7
- 93 systems from 34 teams

Lexical sample task

- Select a sample of words from the lexicon
- Systems must then tag several instances of the sample words in short extracts of text
- SENSEVAL-1: 35 words, 41 tasks
  - 700001 John Dos Passos wrote a poem that talked of `the <tag>bitter</> beat look, the scorn on the lip."
  - 700002 The beans almost double in size during roasting. Black beans are over roasted and will have a <tag>bitter</> flavour and insufficiently roasted beans are pale and give a colourless, tasteless drink.

Lexical sample task: SENSEVAL-1

Nouns		Verbs		Adjectives		Indeterminates	
-n	N	**-v**	N	**-a**	N	**-p**	N
accident	267	amaze	70	brilliant	229	band	302
behaviour	279	bet	177	deaf	122	bitter	373
bet	274	bother	209	floating	47	hurdle	323
disability	160	bury	201	generous	227	sanction	431
excess	186	calculate	217	giant	97	shake	356
float	75	consume	186	modest	270		
giant	118	derive	216	slight	218		
…	…	…	…	…	…		
TOTAL	2756	TOTAL	2501	TOTAL	1406	TOTAL	1785

All-words task

- Systems must tag almost all of the content words in a sample of running text
  - sense-tag all predicates, nouns that are heads of noun-phrase arguments to those predicates, and adjectives modifying those nouns
  - ~5,000 running words of text
  - ~2,000 sense-tagged words

SENSEVAL-2 results

Language	Task	No. of submissions	No. of teams	IAA	Baseline	Best system
Czech	AW	1	1	-	-	.94
Basque	LS	3	2	.75	.65	.76
Estonian	AW	2	2	.72	.85	.67
Italian	LS	2	2	-	-	.39
Korean	LS	2	2	-	.71	.74
Spanish	LS	12	5	.64	.48	.65
Swedish	LS	8	5	.95	-	.70
Japanese	LS	7	3	.86	.72	.78
Japanese	TL	9	8	.81	.37	.79
English	AW	21	12	.75	.57	.69
English	LS	26	15	.86	.51/.16	.64/.40

SENSEVAL plans

- Where next?
  - Supervised ML approaches worked best
    - » Looking the role of feature selection algorithms
  - Need a well-motivated sense inventory
    - » Inter-annotator agreement went down when moving to WordNet senses
  - Need to tie WSD to real applications
    - » The translation task was a good initial attempt