CS6740/INFO6300

- Short intro to word sense disambiguation
  - Lexical semantics
  - Lexical semantic resources: WordNet
  - Word sense disambiguation
    - Supervised machine learning methods
    - WSD evaluation

Introduction to lexical semantics

- Lexical semantics is the study of
  - the systematic meaning-related connections among words and
  - the internal meaning-related structure of each word
- Lexeme
  - an individual entry in the lexicon
  - a pairing of a particular orthographic and phonological form with some form of symbolic meaning representation
- Sense: the lexeme's meaning component
- Lexicon: a finite list of lexemes

Lexical semantic relations: homonymy

- Homonyms: *words that have the same orthographic form and unrelated meanings*
  - Instead, a ***bank***^{1} can hold the investments in a custodial account in the client's name.
  - But as agriculture burgeons on the east ***bank***^{2}, the river will shrink even more.

Lexical semantic relations: polysemy

- Polysemy: the phenomenon of multiple *related* meanings within a single lexeme
  - Example: While some ***banks*** furnish blood only to hospitals, others are much less restrictive.
  - New sense, e.g. bank^{3}?
  - Polysemy allows us to associate a lexeme with a set of related senses.
- Distinguishing homonymy from polysemy is not always easy. Decision is based on:
  - Etymology: history of the lexemes in question
  - Intuition of native speakers

Polysemous lexemes

- For any given single lexeme we would like to be able to answer the following questions:
  - What distinct senses does it have?
    - generally rely on lexicographers
  - How are these senses related?
    - relatively little work in this area
  - How can they be reliably distinguished?
    - this is the task of **word sense disambiguation**

Word sense disambiguation

- Given a *fixed* set of senses 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

CS6740/INFO6300

- Short intro to word sense disambiguation
  - Lexical semantics
  - Lexical semantic resources: WordNet
  - Word sense disambiguation
    - Supervised machine learning methods
    - WSD evaluation

WordNet

- Handcrafted database of lexical relations
- Three separate databases: nouns; verbs; adjectives and adverbs
- Each database is a set of lexical entries (unique orthographic forms)
  - Entries described and indexed in terms of *synsets,* i.e., sets of synonyms (lexemes with the “same meaning”)

Sample WordNet entry

The noun “bass” has 8 senses in WordNet.

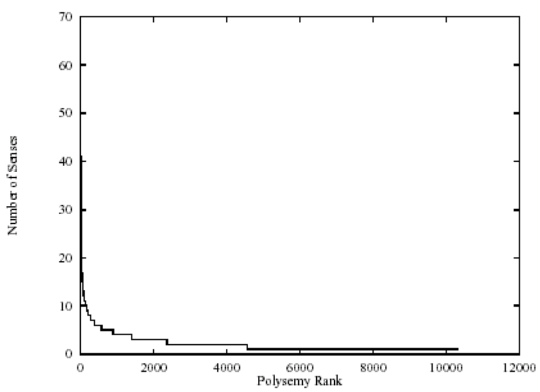
1. bass - (the lowest part of the musical range)
2. bass, bass part - (the lowest part in polyphonic music)
3. bass, basso - (an adult male singer with the lowest voice)
4. sea bass, bass - (flesh of lean-fleshed saltwater fish of the family Serranidae)
5. freshwater bass, bass - (any of various North 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)

Some WordNet Statistics

Part-of-speech	Avg Polysemy	Avg Polysemy w/o monosemous words
Noun	1.24	2.79
Verb	2.17	3.57
Adjective	1.40	2.71
Adverb	1.25	2.50

Distribution of senses

- Zipf distribution of senses



WordNet relations (among **synsets**)

- 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*

CS6740/INFO6300

- Short intro to word sense disambiguation
  - Lexical semantics
  - Lexical semantic resources: WordNet
  - Word sense disambiguation
    - » Supervised machine learning methods
    - » WSD evaluation

Word sense disambiguation

- Given a *fixed* set of senses associated with a lexical item, determine which of them applies to a particular instance of the lexical item

*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.*

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)

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 in context C
    - » Retrieve all of the sense definitions, *S*, for *W* from the MRD
    - » Compare each *s* in *S* to the dictionary definitions D of all the remaining words c in the context C
    - » Select the sense *s* with the most overlap with D (the definitions of the context words C)

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.

Supervised Inductive ML framework

Examples of task

(features + class)

description of context

correct word sense

ML Algorithm

Novel example

(features)

Classifier

(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, boolean, categorical, …

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 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.*

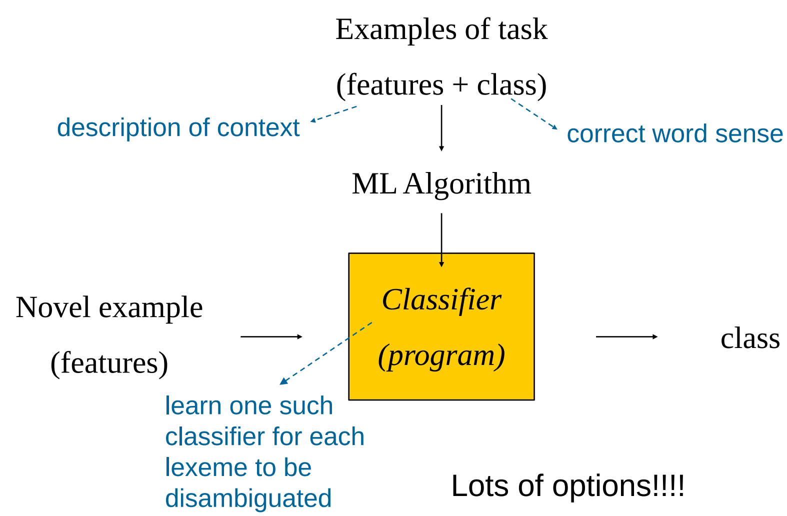
pre2-word	pre2-pos	pre1-word	pre1-pos	fol1-word	fol1-pos	fol2-word	fol2-pos
guitar	NN1	and	CJC	player	NN1	stand	VVB

Co-occurrence features

- Encodes information about neighboring words, ignoring exact positions.
  - 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)
  - **Attributes**: the words themselves (or their roots)
  - **Values**: number of times the word occurs in a region surrounding the target word

fishing?	big?	sound?	player?	fly?	rod?	pound?	double?	…	guitar?	band?
0	0	0	1	0	0	0	0		1	0

Inductive ML framework



Lots of options!!!!

CS6740/INFO6300

- Short intro to word sense disambiguation
  - Lexical semantics
  - Lexical semantic resources: WordNet
  - Word sense disambiguation
    - » Supervised machine learning methods
    - » WSD evaluation

SENSEVAL

- Three tasks (originally)
  - Lexical sample
  - All-words
  - Translation
- Multiple (12+) languages
- Lexicon
  - SENSEVAL-1: from HECTOR corpus
  - SENSEVAL-2: from WordNet 1.7
- Lots of community participation
  - SENSEVAL-1 (1998): 93 systems from 34 teams

Lexical sample task

- Select a sample of words from the lexicon
- Systems must then tag instances of the sample words in short extracts of text
- SENSEVAL-1: 35 words
  - 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

Translation task

- SENSEVAL-2 task
- Only for Japanese
- word sense is defined according to translation distinction
  - if the head word is translated differently in the given expressional context, then it is treated as constituting a different sense
- word sense disambiguation involves selecting the appropriate English word/phrase/sentence equivalent for a Japanese word

SENSEVAL-2 results (2001)

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-2 de-briefing

- Where next?
  - Supervised ML approaches worked best
    - » Looking at the role of feature selection algorithms
  - Need a well-motivated sense inventory
    - » Inter-annotator agreement went down vs. SENSEVAL-1 (moved to WordNet senses)
  - Need to tie WSD to real applications
    - » The translation task was a good initial attempt

SENSEVAL-3 2004

- 14 core WSD tasks including
  - All words (Eng, Italian): 5000 word sample
  - Lexical sample (7 languages)
- Tasks for identifying semantic roles, for multilingual annotations, logical form, subcategorization frame acquisition
- Evaluations now called SEMEVAL

Results

- 27 teams, 47 systems
- Most frequent sense baseline
  - 55.2% (fine-grained)
  - 64.5% (coarse)
- Most systems significantly above baseline
  - Including some unsupervised systems
- Best system
  - 72.9% (fine-grained)
  - 79.3% (coarse)

SENSEVAL-3 lexical sample results

System/Team	Description	Fine P	Fine R	Coarse P	Coarse R
htsa3 U.Bucharest (Grozea)	A Naive Bayes system, with correction of the a-priori frequencies, by dividing the output confidence of the senses by $frequency^{\alpha}$ ($\alpha = 0.2$)	72.9	72.9	79.3	79.3
IRST-Kernels ITC-IRST (Strapparava)	Kernel methods for pattern abstraction, paradigmatic and syntagmatic info. and unsupervised term proximity (LSA) on BNC, in an SVM classifier.	72.6	72.6	79.5	79.5
nusels Nat.U. Singapore (Lee)	A combination of knowledge sources (part-of-speech of neighbouring words, words in context, local collocations, syntactic relations), in an SVM classifier.	72.4	72.4	78.8	78.8
htsa4	Similar to htsa3, with different correction function of a-priori frequencies.	72.4	72.4	78.8	78.8
BCU_comb Basque Country U. (Agirre & Martinez)	An ensemble of decision lists, SVM, and vectorial similarity, improved with a variety of smoothing techniques. The features consist of local collocations, syntactic dependencies, bag-of-words, domain features.	72.3	72.3	78.9	78.9
htsa1	Similar to htsa3, but with smaller number of features.	72.2	72.2	78.7	78.7
rlsc-comb U.Bucharest (Popescu)	A regularized least-square classification (RLSC), using local and topical features, with a term weighting scheme.	72.2	72.2	78.4	78.4
htsa2	Similar to htsa4, but with smaller number of features.	72.1	72.1	78.6	78.6
BCU_english	Similar to BCU_comb, but with a vectorial space model learning.	72.0	72.0	79.1	79.1

SENSEVAL-3 results (unsupervised)

System/Team	Description	Fine P	Fine R	Coarse P	Coarse R
wsdiit IIT Bombay (Ramakrishnan et al.)	An unsupervised system using a Lesk-like similarity between context of ambiguous words, and dictionary definitions. Experiments are performed for various window sizes, various similarity measures	66.1	65.7	73.9	74.1
Cymfony (Niu)	A Maximum Entropy model for unsupervised clustering, using neighboring words and syntactic structures as features. A few annotated instances are used to map context clusters to WordNet/Worsmyth senses.	56.3	56.3	66.4	66.4
Prob0 Cambridge U. (Preiss)	A combination of two unsupervised modules, using basic part of speech and frequency information.	54.7	54.7	63.6	63.6
clr04-ls CL Research (Litkowski)	An unsupervised system relying on definition properties (syntactic, semantic, subcategorization patterns, other lexical information), as given in a dictionary. Performance is generally a function of how well senses are distinguished.	45.0	45.0	55.5	55.5
CIAOSENSO U. Genova (Buscaldi)	An unsupervised system that combines the conceptual density idea with the frequency of words to disambiguate; information about domains is also taken into account.	50.1	41.7	59.1	49.3