CS4740 Natural Language Processing

- Last classes
  - Intro to lexical semantics
  - Lexical semantic resources: WordNet

- Next
  - Word sense disambiguation
    - Dictionary-based approaches
    - Supervised machine learning methods
    - WSD evaluation
    - Weakly supervised methods

Machine learning approaches

- Machine learning paradigms for WSD
  - *Supervised inductive* learning
    - classification
  - Bootstrapping
  - Unsupervised
- Emphasis is on acquiring the knowledge needed for the task from data, rather than from human analysts (e.g., via a set of rules) or from a static algorithm (e.g., Lesk approach)

Supervised 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

- Designed with respect to the **target**, i.e. the word to be disambiguated

- Encodes information from the **context** --- portion of the surrounding text --- that might be useful for determining the word sense of the target word
  - Select a "window" size
  - Extract *features* (also called *attribute-value pairs*) from the context (and possibly the target)
    - Values can be numeric, boolean, categorical, … any type permitted by the learning algorithm

What features to use?

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.*

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

- Encode information about neighboring words, ignoring exact positions.
  - **Attributes**: words highly associated with exactly one of the senses
  - **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 **attributes (features)**
    - *n* 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*
    - window of size 10

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

Labeled training 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.*

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

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

: *music*

guitar, NN1, and, CJC, player, NN1,stand, VVB, 0, 0, 0, 1, 0, ..., 1, 0 : music

Clickers, please

Marseille is annoying when he **begs** for his dinner.

Consider collocational features (window of 4). Which might be the most useful for determining the correct sense of **beg** for the above example?

A. preceding word
B. following word
C. preceding part-of-speech
D. none of the above

- S: (v) **beg#1**, implore#1, pray#2 (call upon in supplication; entreat) *"I beg you to stop!"*
- S: (v) solicit#1, **beg#2**, tap#12 (make a solicitation or entreaty for something; request urgently or persistently) *"Henry IV solicited the Pope for a divorce"; "My neighbor keeps soliciting money for different charities"*
- S: (v) **beg#3** (ask to obtain free) *"beg money and food"*
- S: (v) **beg#4** (dodge, avoid answering, or take for granted) *"beg the question"; "beg the point in the discussion"*

Provide three co-occurrence features (window of 200) that might be useful for determining the correct sense of **beg**.

- S: (v) **beg#1**, implore#1, pray#2 (call upon in supplication; entreat) *"I beg you to stop!"*
- S: (v) solicit#1, **beg#2**, tap#12 (make a solicitation or entreaty for something; request urgently or persistently) *"Henry IV solicited the Pope for a divorce"; "My neighbor keeps soliciting money for different charities"*
- S: (v) **beg#3** (ask to obtain free) *"beg money and food"*
- S: (v) **beg#4** (dodge, avoid answering, or take for granted) *"beg the question"; "beg the point in the discussion"*

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

Decision list classifiers

- Decision lists: equivalent to simple case statements.
  - Consists of a sequence of tests to be applied to each input example/vector; returns a word sense.
  - Each test can check the value of one feature
- 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	$\Rightarrow$	**bass**^{1}
striped bass	$\Rightarrow$	**bass**^{1}
guitar within window	$\Rightarrow$	**bass**^{2}
bass player	$\Rightarrow$	**bass**^{2}
piano within window	$\Rightarrow$	**bass**^{2}
tenor within window	$\Rightarrow$	**bass**^{2}
sea bass	$\Rightarrow$	**bass**^{1}
play/V bass	$\Rightarrow$	**bass**^{2}
river within window	$\Rightarrow$	**bass**^{1}
violin within window	$\Rightarrow$	**bass**^{2}
salmon within window	$\Rightarrow$	**bass**^{1}
on bass	$\Rightarrow$	**bass**^{2}
bass are	$\Rightarrow$	**bass**^{1}

Learning decision lists

- Consists of *generating* and *ordering* individual tests based on the characteristics of the training data
- Generation: every attribute-value pair (i.e. feature) in training set 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

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

CS4740 Natural Language Processing

- Last classes
  - Intro to lexical semantics
  - Lexical semantic resources: WordNet
- Next
  - Word sense disambiguation
    - Dictionary-based approaches
    - Supervised machine learning methods
    - WSD evaluation
    - Weakly supervised methods

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)
    - » Performance evaluation conference (every few years)
    - » Provides an evaluation framework (Kilgarriff and Palmer, 2000)
- Baseline: most frequent sense

Metrics

- Precision
  - # correct / # of predictions
- Recall (== accuracy)
  - # correct / # of examples in test set

WSD Evaluation

- Issues with the Metrics
  - 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 2001

- 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 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

- predicates
- nouns that are heads of noun-phrase arguments to those predicates
- adjectives modifying those nouns

The twentieth century author wrote a poem that talked of `the bitter beat look, the scorn on the lip."

Translation task

- SENSEVAL-2 task
- Only for Japanese
- word sense is defined according to translation distinction
  - if the target 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 equivalent for a Japanese word

English → German

WalMart is **open** from 9 to 5.

Ithaca's WalMart **opened** in 2001.

geoffnet

eroffnet

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-3 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