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

Word sense disambiguation

- Given a *fixed* set of senses associated with a lexical item (usually an orthographic form rather than a lexeme),

- determine which sense applies to a particular instance of the lexical item in context (i.e., in running text)

Marseille found a smelly **bass**
along the **bank** of the Seine.

Example

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

Verb

- S: (v) **beg#1**, implore#1, pray#2 (call upon in supplication; entreat)
- S: (v) solicit#1, **beg#2**, tap#12 (make a solicitation or entreaty for something; request urgently or persistently)
- S: (v) **beg#3** (ask to obtain free)
- S: (v) **beg#4** (dodge, avoid answering, or take for granted)

Example

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

- 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"*

Word sense disambiguation

- 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 in context C
    - » Retrieve all of the sense definitions for *W*, S_W, from the MRD
    - » Compare each *s* in S_W to D_C --- all of the dictionary definitions of all words in C
    - » Select the sense *s* with the largest content-word overlap with D_C

Example

- W = *cone*
- C = { *pine* } **pine cone**
- Sense definitions from MRD

pine	1 kind of evergreen tree with needle-shaped leaves	D_C
	2 waste away through sorrow or illness	
cone	1 solid body which narrows to a point	S_W
	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.

CS4740 Natural Language Processing

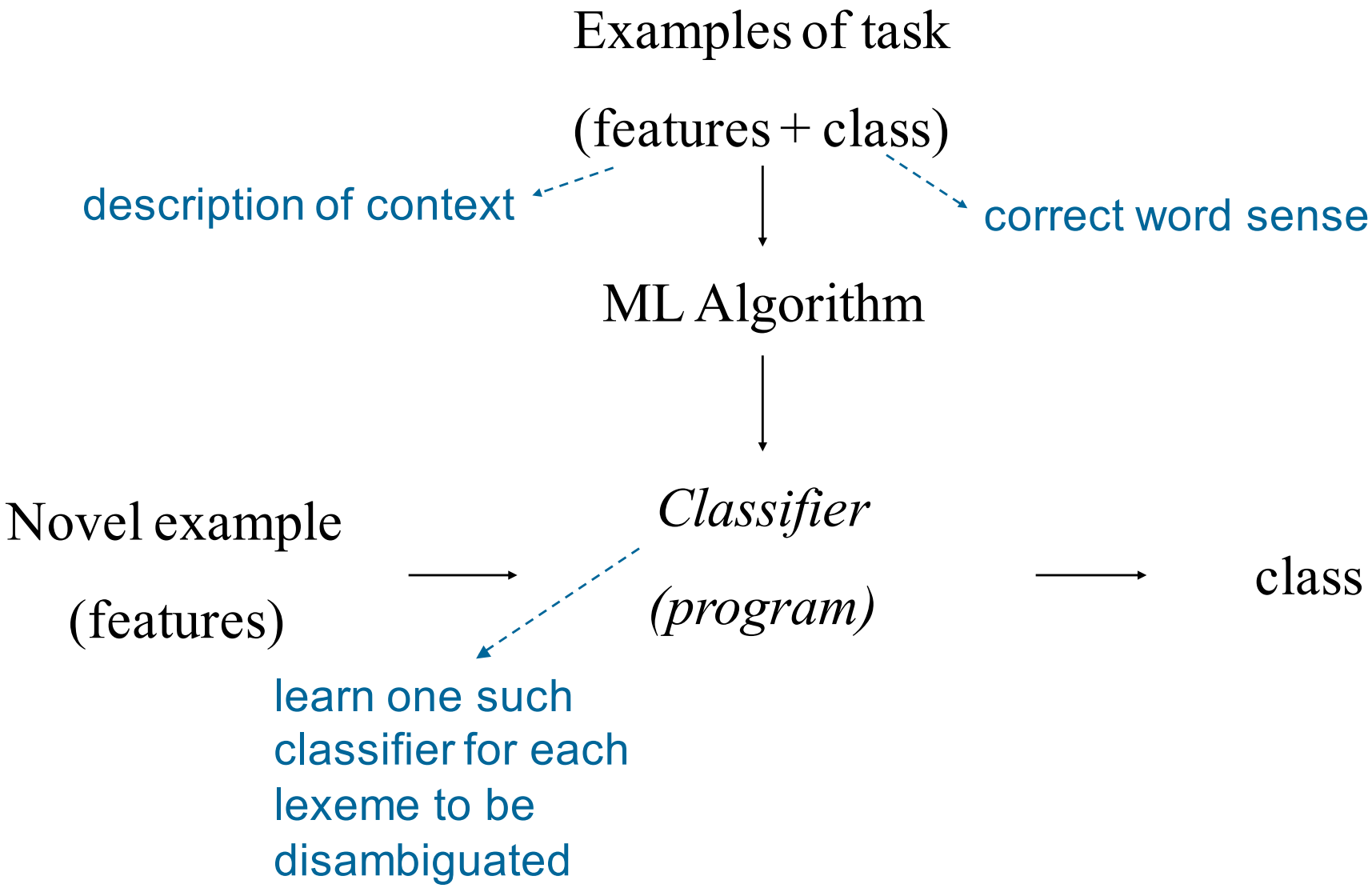
- Last classes
  - Intro to lexical semantics
  - Lexical semantic resources: WordNet

- 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



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

- W.r.t. the **target**, i.e. the word to be disambiguated
- Describe **context** : portion of the surrounding text
  - Select a "window" size
  - Extract features from the context (and possibly the target)
    - Attribute-value pairs
    - Values can be numeric, boolean, categorical, …

What features to use?

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

- 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"*

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.*

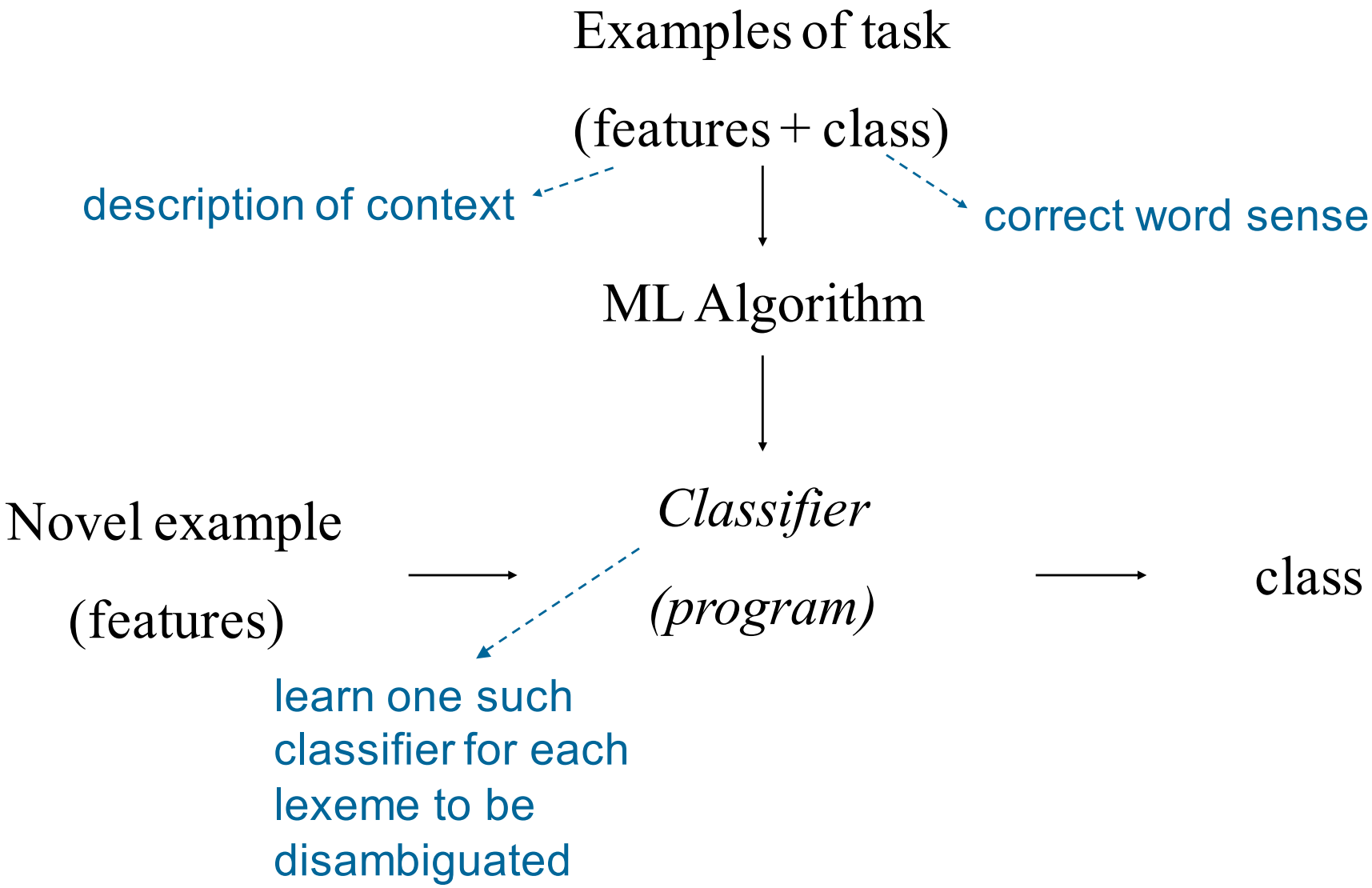
pre2-word	pre2-pos	pre1-word	pre1-pos	fol1-word	fol1-pos	fol2-word	fol2-pos
guitar	NN1	and	CJC	player	NN1	stand	VVB

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

Inductive ML framework



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.
  - 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	⇒	**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 attribute-value pair (i.e. feature) 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