Lexical Semantics & Word Sense Disambiguation

Ling571
Deep Processing Techniques for NLP
February 16, 2011

Roadmap

- Lexical semantics
  - Lexical taxonomy
    - WordNet

  - Thematic Roles
    - Issues
    - Resources:
      - PropBank & FrameNet

  - Selectional Restrictions

  - Primitive decompositions

WordNet Taxonomy

- Most widely used English sense resource
- Manually constructed lexical database
  - 3 Tree-structured hierarchies
    - Nouns (117K) , verbs (11K), adjective+adverb (27K)
    - Entries: synonym set, gloss, example use
- Relations between entries:
  - Synonymy: in synset
  - Hypo(per)nym: Isa tree

WordNet

The noun "bass" has 8 senses in WordNet.

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

The adjective "bass" has 1 sense in WordNet.

1. bass^{1}, deep^{6} - (having or denoting a low vocal or instrumental range) *"a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"*

Noun WordNet Relations

Relation	**Also Called**	**Definition**	**Example**
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Instance Hypernym	Instance	From instances to their concepts	$Austen^1 \rightarrow author^1$
Instance Hyponym	Has-Instance	From concepts to concept instances	$composer^1 \rightarrow Bach^1$
Member Meronym	Has-Member	From groups to their members	$faculty^2 \rightarrow professor^1$
Member Holonym	Member-Of	From members to their groups	$copilot^1 \rightarrow crew^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Substance Meronym		From substances to their subparts	$water^1 \rightarrow oxygen^1$
Substance Holonym		From parts of substances to wholes	$gin^1 \rightarrow martini^1$
Antonym		Semantic opposition between lemmas	$leader^1 \Longleftrightarrow follower^1$
Derivationally Related Form		Lemmas w/same morphological root	$destruction^1 \Longleftrightarrow destroy^1$

WordNet Taxonomy

```
Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist, vocalizer, vocaliser
   => musician, instrumentalist, player
      => performer, performing artist
         => entertainer
            => person, individual, someone...
               => organism, being
                  => living thing, animate thing,
                     => whole, unit
                        => object, physical object
                           => physical entity
                              => entity
               => causal agent, cause, causal agency
                  => physical entity
                     => entity
```

Thematic Roles

- Describe semantic roles of verbal arguments
  - Capture commonality across verbs

Thematic Roles

- Describe semantic roles of verbal arguments
  - Capture commonality across verbs
  - E.g. subject of break, open is AGENT
    - AGENT: volitional cause
    - THEME: things affected by action

Thematic Roles

- Describe semantic roles of verbal arguments
  - Capture commonality across verbs
  - E.g. subject of break, open is AGENT
    - AGENT: volitional cause
    - THEME: things affected by action
  - Enables generalization over surface order of arguments
    - $\text{John}_{\text{AGENT}}$ broke the $\text{window}_{\text{THEME}}$

Thematic Roles

- Describe semantic roles of verbal arguments
  - Capture commonality across verbs
  - E.g. subject of break, open is AGENT
    - AGENT: volitional cause
    - THEME: things affected by action

  - Enables generalization over surface order of arguments
    - $\text{John}_{\text{AGENT}}$ broke the $\text{window}_{\text{THEME}}$
    - The $\text{rock}_{\text{INSTRUMENT}}$ broke the $\text{window}_{\text{THEME}}$

Thematic Roles

- Describe semantic roles of verbal arguments
  - Capture commonality across verbs
  - E.g. subject of break, open is AGENT
    - AGENT: volitional cause
    - THEME: things affected by action
  - Enables generalization over surface order of arguments
    - John$_{AGENT}$ broke the window$_{THEME}$
    - The rock$_{INSTRUMENT}$ broke the window$_{THEME}$
    - The window$_{THEME}$ was broken by John$_{AGENT}$

Thematic Roles

- Thematic grid, θ -grid, case frame
  - Set of thematic role arguments of verb

Thematic Roles

- Thematic grid, θ-grid, case frame
  - Set of thematic role arguments of verb
    - E.g. Subject:AGENT; Object:THEME, or
    - Subject: INSTR; Object: THEME

Thematic Roles

- Thematic grid, θ-grid, case frame
  - Set of thematic role arguments of verb
    - E.g. Subject:AGENT; Object:THEME, or
    - Subject: INSTR; Object: THEME
- Verb/Diathesis Alternations
  - Verbs allow different surface realizations of roles

Thematic Roles

- Thematic grid, θ -grid, case frame
  - Set of thematic role arguments of verb
    - E.g. Subject:AGENT; Object:THEME, or
    - Subject: INSTR; Object: THEME

- Verb/Diathesis Alternations
  - Verbs allow different surface realizations of roles
    - Doris$_{\text{AGENT}}$ gave the book$_{\text{THEME}}$ to Cary$_{\text{GOAL}}$

Thematic Roles

- Thematic grid, θ -grid, case frame
  - Set of thematic role arguments of verb
    - E.g. Subject:AGENT; Object:THEME, or
    - Subject: INSTR; Object: THEME

- Verb/Diathesis Alternations
  - Verbs allow different surface realizations of roles
    - Doris$_{\text{AGENT}}$ gave the book$_{\text{THEME}}$ to Cary$_{\text{GOAL}}$
    - Doris$_{\text{AGENT}}$ gave Cary$_{\text{GOAL}}$ the book$_{\text{THEME}}$

Thematic Roles

- Thematic grid, θ-grid, case frame
  - Set of thematic role arguments of verb
    - E.g. Subject:AGENT; Object:THEME, or
    - Subject: INSTR; Object: THEME

- Verb/Diathesis Alternations
  - Verbs allow different surface realizations of roles
    - $\text{Doris}_{\text{AGENT}}$ gave the $\text{book}_{\text{THEME}}$ to $\text{Cary}_{\text{GOAL}}$
    - $\text{Doris}_{\text{AGENT}}$ gave $\text{Cary}_{\text{GOAL}}$ the $\text{book}_{\text{THEME}}$
  - Group verbs into classes based on shared patterns

Canonical Roles

Thematic Role	**Example**
AGENT	*The waiter* spilled the soup.
EXPERIENCER	*John* has a headache.
FORCE	*The wind* blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke *the ice*...
RESULT	The French government has built a *regulation-size baseball diamond*...
CONTENT	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	He turned to poaching catfish, stunning them *with a shocking device*...
BENEFICIARY	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	I flew in *from Boston*.
GOAL	I drove *to Portland*.

Thematic Role Issues

- Hard to produce

Thematic Role Issues

- Hard to produce
  - Standard set of roles
    - Fragmentation: Often need to make more specific
      - E,g, INSTRUMENTS can be subject or not

Thematic Role Issues

- Hard to produce
  - Standard set of roles
    - Fragmentation: Often need to make more specific
      - E,g, INSTRUMENTS can be subject or not
  - Standard definition of roles
    - Most AGENTs: animate, volitional, sentient, causal
    - But not all....
- Strategies:
  - Generalized semantic roles: PROTO-AGENT/PROTO-PATIENT
    - Defined heuristically: PropBank
  - Define roles specific to verbs/nouns: FrameNet

Thematic Role Issues

- Hard to produce
  - Standard set of roles
    - Fragmentation: Often need to make more specific
      - E,g, INSTRUMENTS can be subject or not
  - Standard definition of roles
    - Most AGENTs: animate, volitional, sentient, causal
    - But not all....

Thematic Role Issues

- Hard to produce
  - Standard set of roles
    - Fragmentation: Often need to make more specific
      - E,g, INSTRUMENTS can be subject or not
  - Standard definition of roles
    - Most AGENTs: animate, volitional, sentient, causal
    - But not all....
- Strategies:
  - Generalized semantic roles: PROTO-AGENT/PROTO-PATIENT
    - Defined heuristically: PropBank

Thematic Role Issues

- Hard to produce
  - Standard set of roles
    - Fragmentation: Often need to make more specific
      - E,g, INSTRUMENTS can be subject or not
  - Standard definition of roles
    - Most AGENTs: animate, volitional, sentient, causal
    - But not all....
- Strategies:
  - Generalized semantic roles: PROTO-AGENT/PROTO-PATIENT
    - Defined heuristically: PropBank
  - Define roles specific to verbs/nouns: FrameNet

PropBank

- Sentences annotated with semantic roles
  - Penn and Chinese Treebank

PropBank

- Sentences annotated with semantic roles
  - Penn and Chinese Treebank
  - Roles specific to verb sense
    - Numbered: Arg0, Arg1, Arg2,...
      - Arg0: PROTO-AGENT; Arg1: PROTO-PATIENT, etc

PropBank

- Sentences annotated with semantic roles
  - Penn and Chinese Treebank
  - Roles specific to verb sense
    - Numbered: Arg0, Arg1, Arg2,...
      - Arg0: PROTO-AGENT; Arg1: PROTO-PATIENT, etc
  - E.g. agree.01
    - Arg0: Agreer

PropBank

- Sentences annotated with semantic roles
  - Penn and Chinese Treebank
  - Roles specific to verb sense
    - Numbered: Arg0, Arg1, Arg2,...
      - Arg0: PROTO-AGENT; Arg1: PROTO-PATIENT, etc
  - E.g. agree.01
    - Arg0: Agreer
    - Arg1: Proposition

PropBank

- Sentences annotated with semantic roles
  - Penn and Chinese Treebank
  - Roles specific to verb sense
    - Numbered: Arg0, Arg1, Arg2,...
      - Arg0: PROTO-AGENT; Arg1: PROTO-PATIENT, etc
  - E.g. agree.01
    - Arg0: Agreer
    - Arg1: Proposition
    - Arg2: Other entity agreeing

PropBank

- Sentences annotated with semantic roles
  - Penn and Chinese Treebank
  - Roles specific to verb sense
    - Numbered: Arg0, Arg1, Arg2,...
      - Arg0: PROTO-AGENT; Arg1: PROTO-PATIENT, etc
  - E.g. agree.01
    - Arg0: Agreer
    - Arg1: Proposition
    - Arg2: Other entity agreeing
    - Ex1: [$_{Arg0}$The group] agreed [$_{Arg1}$it wouldn't make an offer]

FrameNet

- Semantic roles specific to Frame
  - Frame: script-like structure, roles (frame elements)

FrameNet

- Semantic roles specific to Frame
  - Frame: script-like structure, roles (frame elements)

  - E.g. change_position_on_scale: increase, rise
    - Attribute, Initial_value, Final_value

FrameNet

- Semantic roles specific to Frame
  - Frame: script-like structure, roles (frame elements)

  - E.g. change_position_on_scale: increase, rise
    - Attribute, Initial_value, Final_value

  - Core, non-core roles

FrameNet

- Semantic roles specific to Frame
  - Frame: script-like structure, roles (frame elements)
  - E.g. change_position_on_scale: increase, rise
    - Attribute, Initial_value, Final_value
  - Core, non-core roles
  - Relationships b/t frames, frame elements
    - Add causative: cause_change_position_on_scale

Core Roles	
ATTRIBUTE	The ATTRIBUTE is a scalar property that the ITEM possesses.
DIFFERENCE	The distance by which an ITEM changes its position on the scale.
FINAL_STATE	A description that presents the ITEM's state after the change in the ATTRIBUTE's value as an independent predication.
FINAL_VALUE	The position on the scale where the ITEM ends up.
INITIAL_STATE	A description that presents the ITEM's state before the change in the ATTRIBUTE's value as an independent predication.
INITIAL_VALUE	The initial position on the scale from which the ITEM moves away.
ITEM	The entity that has a position on the scale.
VALUE_RANGE	A portion of the scale, typically identified by its end points, along which the values of the ATTRIBUTE fluctuate.
Some Non-Core Roles	
DURATION	The length of time over which the change takes place.
SPEED	The rate of change of the VALUE.
GROUP	The GROUP in which an ITEM changes the value of an ATTRIBUTE in a specified way.

Selectional Restrictions

- Semantic type constraint on arguments
  - *I want to eat someplace close to UW*

Selectional Restrictions

- Semantic type constraint on arguments
  - *I want to eat someplace close to UW*
  - E.g. THEME of eating should be edible
    - Associated with senses

Selectional Restrictions

- Semantic type constraint on arguments
  - *I want to eat someplace close to UW*
  - E.g. THEME of eating should be edible
    - Associated with senses
- Vary in specificity:

Selectional Restrictions

- Semantic type constraint on arguments
  - *I want to eat someplace close to UW*
  - E.g. THEME of eating should be edible
    - Associated with senses
- Vary in specificity:
  - Imagine: AGENT: human/sentient; THEME: any

Selectional Restrictions

- Semantic type constraint on arguments
  - *I want to eat someplace close to UW*
  - E.g. THEME of eating should be edible
    - Associated with senses
- Vary in specificity:
  - Imagine: AGENT: human/sentient; THEME: any
- Representation:
  - Add as predicate in FOL event representation

Selectional Restrictions

- Semantic type constraint on arguments
  - *I want to eat someplace close to UW*
  - E.g. THEME of eating should be edible
    - Associated with senses
- Vary in specificity:
  - Imagine: AGENT: human/sentient; THEME: any
- Representation:
  - Add as predicate in FOL event representation
    - Overkill computationally; requires large commonsense KB

Selectional Restrictions

- Semantic type constraint on arguments
  - *I want to eat someplace close to UW*
  - E.g. THEME of eating should be edible
    - Associated with senses
- Vary in specificity:
  - Imagine: AGENT: human/sentient; THEME: any
- Representation:
  - Add as predicate in FOL event representation
    - Overkill computationally; requires large commonsense KB
  - Associate with WordNet synset (and hyponyms)

Primitive Decompositions

- Jackendoff(1990), Dorr(1999), McCawley (1968)

- Word meaning constructed from primitives
  - Fixed small set of basic primitives
    - E.g. cause, go, become,
    - kill=cause X to become Y

Primitive Decompositions

- Jackendoff(1990), Dorr(1999), McCawley (1968)

- Word meaning constructed from primitives
  - Fixed small set of basic primitives
    - E.g. cause, go, become,
    - kill=cause X to become Y
  - Augment with open-ended “manner”
    - Y = not alive
    - E.g. walk vs run

Primitive Decompositions

- Jackendoff(1990), Dorr(1999), McCawley (1968)

- Word meaning constructed from primitives
  - Fixed small set of basic primitives
    - E.g. cause, go, become,
    - kill=cause X to become Y
  - Augment with open-ended "manner"
    - Y = not alive
    - E.g. walk vs run
- Fixed primitives/Infinite descriptors

Word Sense Disambiguation

- Selectional Restriction-based approaches
  - Limitations
- Robust Approaches
  - Supervised Learning Approaches
    - Naïve Bayes
  - Dictionary-based Approaches
  - Bootstrapping Approaches
    - One sense per discourse/collocation
  - Unsupervised Approaches
    - Schutze's word space
  - Resource-based Approaches
    - Dictionary parsing, WordNet Distance
  - Why they work
  - Why they don't

Word Sense Disambiguation

- Application of lexical semantics
- Goal: Given a word *in context,* identify the appropriate sense
  - E.g. <u>plants</u> and animals in the rainforest
- Crucial for real syntactic & semantic analysis

Word Sense Disambiguation

- Application of lexical semantics
- Goal: Given a word *in context,* identify the appropriate sense
  - E.g. plants and animals in the rainforest
- Crucial for real syntactic & semantic analysis
  - Correct sense can determine
    - .

Word Sense Disambiguation

- Application of lexical semantics
- Goal: Given a word *in context,* identify the appropriate sense
  - E.g. plants and animals in the rainforest
- Crucial for real syntactic & semantic analysis
  - Correct sense can determine
    - Available syntactic structure
    - Available thematic roles, correct meaning,..

Selectional Restriction Approaches

- Integrate sense selection in parsing and semantic analysis – e.g. with lambda calculus
- Concept: Predicate selects sense
  - Washing dishes vs stir-frying dishes

Selectional Restriction Approaches

- Integrate sense selection in parsing and semantic analysis – e.g. with lambda calculus
- Concept: Predicate selects sense
  - Washing dishes vs stir-frying dishes
    - Stir-fry: patient: food => dish~food
  -

Selectional Restriction Approaches

- Integrate sense selection in parsing and semantic analysis – e.g. with lambda calculus
- Concept: Predicate selects sense
  - Washing dishes vs stir-frying dishes
    - Stir-fry: patient: food => dish~food
  - Serve Denver vs serve breakfast
    - Serve vegetarian dishes

Selectional Restriction Approaches

- Integrate sense selection in parsing and semantic analysis – e.g. with lambda calculus
- Concept: Predicate selects sense
  - Washing dishes vs stir-frying dishes
    - Stir-fry: patient: food => dish~food
  - Serve Denver vs serve breakfast
    - Serve vegetarian dishes
      - Serve1: patient: loc; serve1: patient: food
      - => dishes~food: only valid variant

Selectional Restriction Approaches

- Integrate sense selection in parsing and semantic analysis – e.g. with lambda calculus
- Concept: Predicate selects sense
  - Washing dishes vs stir-frying dishes
    - Stir-fry: patient: food => dish~food
  - Serve Denver vs serve breakfast
    - Serve vegetarian dishes
      - Serve1: patient: loc; serve1: patient: food
      - => dishes~food: only valid variant
- Integrate in rule-to-rule: test e.g. in WN

Selectional Restrictions: Limitations

- Problem 1:

Selectional Restrictions: Limitations

- Problem 1: Predicates too general
  - Recommend, like, hit....
- Problem 2:

Selectional Restrictions: Limitations

- Problem 1: Predicates too general
  - Recommend, like, hit....
- Problem 2: Language too flexible
  - "The circus performer ate fire and swallowed swords"
    - Unlikely but doable
  - Also metaphor

Selectional Restrictions: Limitations

- Problem 1: Predicates too general
  - Recommend, like, hit….
- Problem 2: Language too flexible
  - "The circus performer ate fire and swallowed swords"
    - Unlikely but doable
  - Also metaphor
- Strong restrictions would block all analysis

Selectional Restrictions: Limitations

- Problem 1: Predicates too general
  - Recommend, like, hit....
- Problem 2: Language too flexible
  - "The circus performer ate fire and swallowed swords"
    - Unlikely but doable
  - Also metaphor
- Strong restrictions would block all analysis
  - Some approaches generalize up hierarchy
    - Can over-accept truly weird things

Selectional Restrictions: Limitations

- Problem 1: Predicates too general
  - Recommend, like, hit….
- Problem 2: Language too flexible
  - "The circus performer ate fire and swallowed swords"
    - Unlikely but doable
  - Also metaphor
- Strong restrictions would block all analysis
  - Some approaches generalize up hierarchy
    - Can over-accept truly weird things
  - Selectional preferences: apply weighted preferences

Robust Disambiguation

- More to semantics than P-A structure
  - Select sense where predicates underconstrain

Robust Disambiguation

- More to semantics than P-A structure
  - Select sense where predicates underconstrain
- Learning approaches
  - Supervised, Bootstrapped, Unsupervised

Robust Disambiguation

- More to semantics than P-A structure
  - Select sense where predicates underconstrain
- Learning approaches
  - Supervised, Bootstrapped, Unsupervised
- Knowledge-based approaches
  - Dictionaries, Taxonomies
- Widen notion of context for sense selection

Robust Disambiguation

- More to semantics than P-A structure
  - Select sense where predicates underconstrain
- Learning approaches
  - Supervised, Bootstrapped, Unsupervised
- Knowledge-based approaches
  - Dictionaries, Taxonomies
- Widen notion of context for sense selection
  - Words within window (2,50,discourse)
  - Narrow cooccurrence - collocations

Disambiguation Features

- Key: What are the features?

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors
  - Morphologically simplified form

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors
  - Morphologically simplified form
  - Words in neighborhood

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors
  - Morphologically simplified form
  - Words in neighborhood
    - Question: How big a neighborhood?

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors
  - Morphologically simplified form
  - Words in neighborhood
    - Question: How big a neighborhood?
      - Is there a single optimal size? Why?
  - ..

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors
  - Morphologically simplified form
  - Words in neighborhood
    - Question: How big a neighborhood?
      - Is there a single optimal size? Why?
  - (Possibly shallow) Syntactic analysis
    - E.g. predicate-argument relations, modification, phrases
  - Collocation vs co-occurrence features

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors
  - Morphologically simplified form
  - Words in neighborhood
    - Question: How big a neighborhood?
      - Is there a single optimal size? Why?
  - (Possibly shallow) Syntactic analysis
    - E.g. predicate-argument relations, modification, phrases
  - Collocation vs co-occurrence features
    - Collocation: words in specific relation: p-a, 1 word +/-

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors
  - Morphologically simplified form
  - Words in neighborhood
    - Question: How big a neighborhood?
      - Is there a single optimal size? Why?
  - (Possibly shallow) Syntactic analysis
    - E.g. predicate-argument relations, modification, phrases
  - Collocation vs co-occurrence features
    - Collocation: words in specific relation: p-a, 1 word +/-
    - Co-occurrence: bag of words..

WSD Evaluation

- Ideally, end-to-end evaluation with WSD component
  - Demonstrate real impact of technique in system

WSD Evaluation

- Ideally, end-to-end evaluation with WSD component
  - Demonstrate real impact of technique in system
  - Difficult, expensive, still application specific

WSD Evaluation

- Ideally, end-to-end evaluation with WSD component
  - Demonstrate real impact of technique in system
  - Difficult, expensive, still application specific
- Typically, intrinsic, sense-based
  - Accuracy, precision, recall
  - SENSEVAL/SEMEVAL: all words, lexical sample
- Baseline:

WSD Evaluation

- Ideally, end-to-end evaluation with WSD component
  - Demonstrate real impact of technique in system
  - Difficult, expensive, still application specific
- Typically, intrinsic, sense-based
  - Accuracy, precision, recall
  - SENSEVAL/SEMEVAL: all words, lexical sample
- Baseline:
  - Most frequent sense, Lesk
- Topline:

WSD Evaluation

- Ideally, end-to-end evaluation with WSD component
  - Demonstrate real impact of technique in system
  - Difficult, expensive, still application specific
- Typically, intrinsic, sense-based
  - Accuracy, precision, recall
  - SENSEVAL/SEMEVAL: all words, lexical sample
- Baseline:
  - Most frequent sense, Lesk
- Topline:
  - Human inter-rater agreement: 75-80% fine; 90% coarse

Naïve Bayes' Approach

- Supervised learning approach
  - Input: feature vector X label

Naïve Bayes' Approach

- Supervised learning approach
  - Input: feature vector X label
- Best sense = most probable sense given f

$$\hat{s} = \arg\max_{s \in S} P(s \mid \vec{f})$$

$$\hat{s} = \arg\max_{s \in S} \frac{P(\vec{f} \mid s) P(s)}{P(\vec{f})}$$

Naïve Bayes' Approach

- Issue:
  - Data sparseness: full feature vector rarely seen

Naïve Bayes' Approach

- Issue:
  - Data sparseness: full feature vector rarely seen
- "Naïve" assumption:
  - Features independent given sense

$$P(\vec{f} \mid s) \approx \prod_{j=1}^{n} P(f_j \mid s)$$

$$\hat{s} = \operatorname*{argmax}_{s \in S} P(s) \prod_{j=1}^{n} P(f_j \mid s)$$

Training NB Classifier

$$\hat{s} = \underset{s \in S}{\operatorname{argmax}} P(s) \prod_{j=1}^{n} P(f_j \mid s)$$

- Estimate P(s):
  - Prior

Training NB Classifier

$$\hat{s} = \underset{s \in S}{\operatorname{argmax}} P(s) \prod_{j=1}^{n} P(f_j \mid s)$$

- Estimate P(s):
  - Prior

$$P(s_i) = \frac{count(s_i, w_j)}{count(w_j)}$$

Training NB Classifier

$$\hat{s} = \underset{s \in S}{\operatorname{argmax}} P(s) \prod_{j=1}^{n} P(f_j \mid s)$$

- Estimate P(s):
  - Prior

$$P(s_i) = \frac{count(s_i, w_j)}{count(w_j)}$$

- Estimate $P(f_j|s)$

Training NB Classifier

$$\hat{s} = \operatorname*{argmax}_{s \in S} P(s) \prod_{j=1}^{n} P(f_j \mid s)$$

- Estimate P(s):
  - Prior

$$P(s_i) = \frac{count(s_i, w_j)}{count(w_j)}$$

- Estimate $P(f_j|s)$

$$P(f_j \mid s) = \frac{count(f_j, s)}{count(s)}$$

- Issues:

Training NB Classifier

$$\hat{s} = \underset{s \in S}{\operatorname{argmax}}\, P(s) \prod_{j=1}^{n} P(f_j \mid s)$$

- Estimate P(s):
  - Prior

$$P(s_i) = \frac{count(s_i, w_j)}{count(w_j)}$$

- Estimate $P(f_j|s)$

$$P(f_j \mid s) = \frac{count(f_j, s)}{count(s)}$$

- Issues:
  - Underflow => log prob

Training NB Classifier

$$\hat{s} = \operatorname*{argmax}_{s \in S} P(s) \prod_{j=1}^{n} P(f_j \mid s)$$

- Estimate P(s):
  - Prior

$$P(s_i) = \frac{count(s_i, w_j)}{count(w_j)}$$

- Estimate $P(f_j|s)$ $$P(f_j \mid s) = \frac{count(f_j, s)}{count(s)}$$
- Issues:
  - Underflow => log prob
  - Sparseness => smoothing

Dictionary-Based Approach

- (Simplified) Lesk algorithm
  - “How to tell a pine cone from an ice cream cone”

Dictionary-Based Approach

- (Simplified) Lesk algorithm
  - “How to tell a pine cone from an ice cream cone”

- Compute context ‘signature’ of word to disambiguate
  - Words in surrounding sentence(s)

Dictionary-Based Approach

- (Simplified) Lesk algorithm
  - "How to tell a pine cone from an ice cream cone"
- Compute context 'signature' of word to disambiguate
  - Words in surrounding sentence(s)
- Compare overlap w.r.t. dictionary entries for senses

Dictionary-Based Approach

- (Simplified) Lesk algorithm
  - “How to tell a pine cone from an ice cream cone”
- Compute context ‘signature’ of word to disambiguate
  - Words in surrounding sentence(s)
- Compare overlap w.r.t. dictionary entries for senses
- Select sense with highest (non-stopword) overlap

Applying Lesk

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

$bank^1$	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”

Applying Lesk

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

$bank^1$	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"

- $Bank^1$: 2

Applying Lesk

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

$bank^1$	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”

- $Bank^1$: 2
- $Bank^2$: 0

Improving Lesk

- Overlap score:
  - All words equally weighted (excluding stopwords)

Improving Lesk

- Overlap score:
  - All words equally weighted (excluding stopwords)

- Not all words equally informative

Improving Lesk

- Overlap score:
  - All words equally weighted (excluding stopwords)

- Not all words equally informative
  - Overlap with unusual/specific words – better
  - Overlap with common/non-specific words – less good

Improving Lesk

- Overlap score:
  - All words equally weighted (excluding stopwords)

- Not all words equally informative
  - Overlap with unusual/specific words – better
  - Overlap with common/non-specific words – less good
- Employ corpus weighting:
  - IDF: inverse document frequency
    - $Idf_i = \log(Ndoc/nd_i)$

Minimally Supervised WSD

  - Yarowsky's algorithm (1995)
- Bootstrapping approach:
  - Use small labeled seedset to iteratively train

Minimally Supervised WSD

- Yarowsky's algorithm (1995)
- Bootstrapping approach:
  - Use small labeled seedset to iteratively train
- Builds on 2 key insights:
  - One Sense Per Discourse
    - Word appearing multiple times in text has same sense
    - Corpus of 37232 bass instances: always single sense

Minimally Supervised WSD

- Yarowsky's algorithm (1995)

- Bootstrapping approach:
  - Use small labeled seedset to iteratively train
- Builds on 2 key insights:
  - One Sense Per Discourse
    - Word appearing multiple times in text has same sense
    - Corpus of 37232 bass instances: always single sense
  - One Sense Per Collocation
    - Local phrases select single sense
      - Fish -> $Bass^1$
      - Play -> $Bass^2$

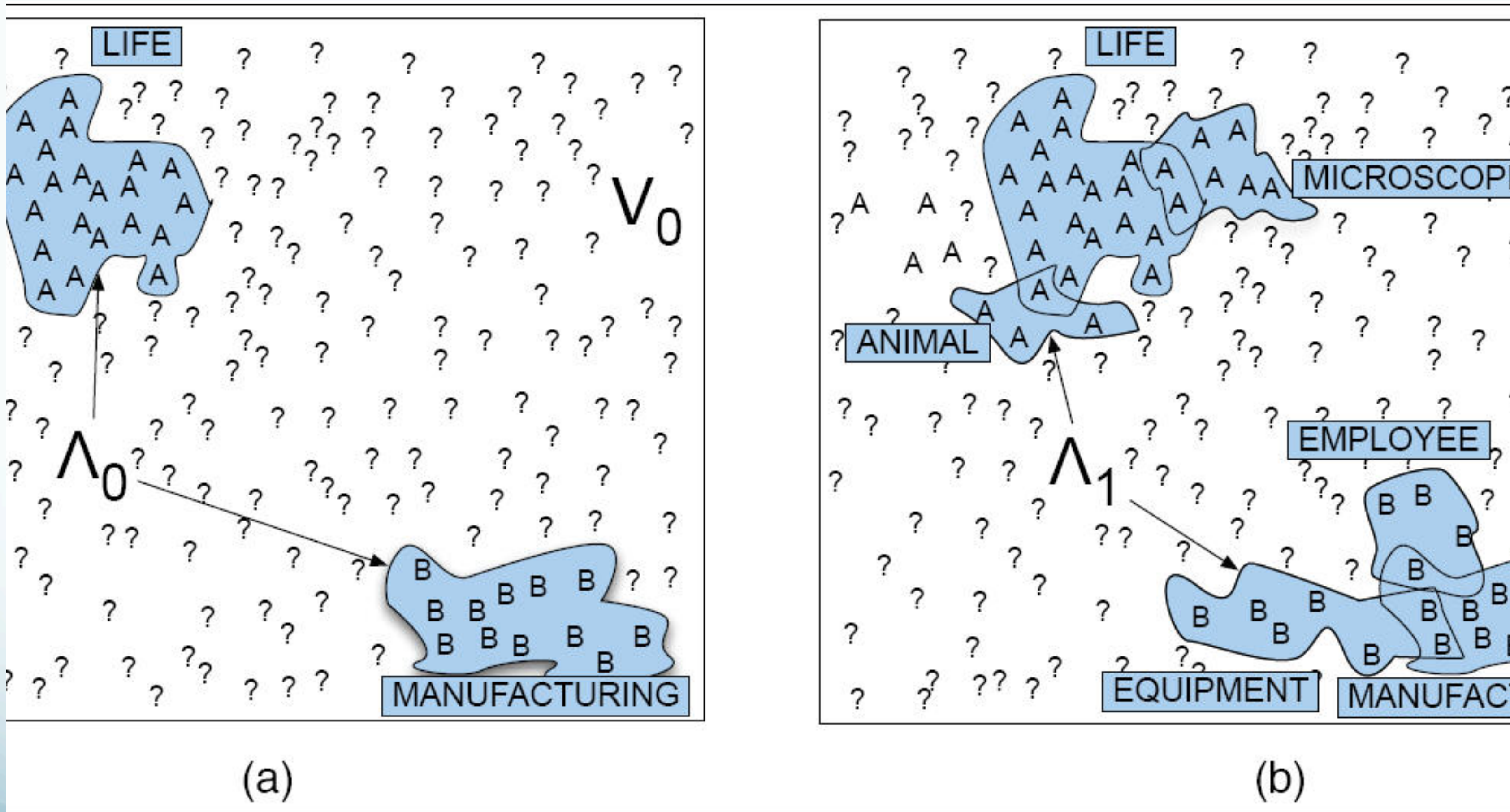
Yarowsky's Algorithm

- Training Decision Lists
  - 1. Pick Seed Instances & Tag
  - 2. Find Collocations: Word Left, Word Right, Word $\pm$K
    - (A) Calculate Informativeness on Tagged Set,
      - Order: $abs(\log \frac{P(Sense_1 | Collocation)}{P(Sense_2 | Collocation)}$
    - (B) Tag New Instances with Rules
    - (C)* Apply 1 Sense/Discourse
    - (D) If Still Unlabeled, Go To 2
  - 3. Apply 1 Sense/Discourse
- Disambiguation: First Rule Matched

Initial decision list for *plant* (abbreviated)

LogL	Collocation	Sense
8.10	*plant* life	⇒ A
7.58	manufacturing *plant*	⇒ B
7.39	life (within ±2-10 words)	⇒ A
7.20	manufacturing (in ±2-10 words)	⇒ B
6.27	animal (within ±2-10 words)	⇒ A
4.70	equipment (within ±2-10 words)	⇒ B
4.39	employee (within ±2-10 words)	⇒ B
4.30	assembly *plant*	⇒ B
4.10	*plant* closure	⇒ B
3.52	*plant* species	⇒ A
3.48	automate (within ±2-10 words)	⇒ B
3.45	microscopic *plant*	⇒ A
	...	

Iterative Updating



There are more kinds of plants and animals in the rainforests than anywhere else on Earth. Over half of the millions of known species of plants and animals live in the rainforest. Many are found nowhere else. There are even plants and animals in the rainforest that we have not yet discovered.

Biological Example

The Paulus company was founded in 1938. Since those days the product range has been the subject of constant expansions and is brought up continuously to correspond with the state of the art. We're engineering, manufacturing and commissioning world-wide ready-to-run plants packed with our comprehensive know-how. Our Product Range includes pneumatic conveying systems for carbon, carbide, sand, lime andmany others. We use reagent injection in molten metal for the…

Industrial Example

Label the First Use of "Plant"

Sense Choice With Collocational Decision Lists

- Create Initial Decision List
  - Rules Ordered by $abs(\log \frac{P(Sense_1 \mid Collocation)}{P(Sense_2 \mid Collocation)}$
- Check nearby Word Groups (Collocations)
  - Biology: "Animal" in $\pm$ 2-10 words
  - Industry: "Manufacturing" in $\pm$ 2-10 words
- Result: Correct Selection
  - 95% on Pair-wise tasks

Schutze' s Vector Space: Detail

- Build a co-occurrence matrix
  - Restrict Vocabulary to 4 letter sequences
  - Exclude Very Frequent - Articles, Affixes
  - Entries in 5000-5000 Matrix

$\rightarrow$ 97 Real Values

- Word Context
  - 4grams within 1001 Characters
  - Sum & Normalize Vectors for each 4gram
  - Distances between Vectors by dot product

Schutze's Vector Space: continued

- Word Sense Disambiguation
  - Context Vectors of All Instances of Word
  - Automatically Cluster Context Vectors
  - Hand-label Clusters with Sense Tag
  - Tag New Instance with Nearest Cluster

Sense Selection in “Word Space”

- Build a Context Vector
  - 1,001 character window - Whole Article
- Compare Vector Distances to Sense Clusters
  - Only 3 Content Words in Common
  - Distant Context Vectors
  - Clusters - Build Automatically, Label Manually
- Result: 2 Different, Correct Senses
  - 92% on Pair-wise tasks

Resnik's WordNet Labeling: Detail

- Assume Source of Clusters
- Assume KB: Word Senses in WordNet IS-A hierarchy
- Assume a Text Corpus
- Calculate Informativeness
  - For Each KB Node:
    - Sum occurrences of it and all children
    - Informativeness

$$(\mathrm{I}) = -\log\left(\frac{\sum_{w \in C} Count(w)}{N}\right)$$

- Disambiguate wrt Cluster & WordNet
  - Find MIS for each pair, I
  - For each subsumed sense, Vote += I
  - Select Sense with Highest Vote

Sense Labeling Under WordNet

- Use Local Content Words as Clusters
  - Biology: Plants, Animals, Rainforests, species...
  - Industry: Company, Products, Range, Systems...
- Find Common Ancestors in WordNet
  - Biology: Plants & Animals isa Living Thing
  - Industry: Product & Plant isa Artifact isa Entity
  - Use Most Informative
- Result: Correct Selection

$$(\mathrm{I}) = -\log\left(\frac{\sum_{w \in C} Count(w)}{N}\right)$$

The Question of Context

- Shared Intuition:
  - Context -> Sense
- Area of Disagreement:
  - What is context?
- Wide vs Narrow Window
- Word Co-occurrences

Taxonomy of Contextual Information

- Topical Content
- Word Associations
- Syntactic Constraints
- Selectional Preferences
- World Knowledge & Inference

A Trivial Definition of Context

All Words within X words of Target

- Many words: Schutze - 1000 characters, several sentences
- Unordered "Bag of Words"
- Information Captured: Topic & Word Association
- Limits on Applicability
  - Nouns vs. Verbs & Adjectives
  - Schutze: Nouns - 92%, "Train" -Verb, 69%

Limits of Wide Context

- Comparison of Wide-Context Techniques (LTV ‘93)
  - Neural Net, Context Vector, Bayesian Classifier, Simulated Annealing
    - Results: 2 Senses - 90+%; 3+ senses ~ 70%
    - People: Sentences ~100%; Bag of Words: ~70%
- Inadequate Context
- Need Narrow Context
  - Local Constraints Override
  - Retain Order, Adjacency

Surface Regularities = Useful Disambiguators

- Not Necessarily!
- Right for the Wrong Reason
  - Burglar Rob… Thieves Stray Crate Chase Lookout
- Learning the Corpus, not the Sense
  - The “Ste.” Cluster: Dry Oyster Whisky Hot Float Ice
- Learning Nothing Useful, Wrong Question
  - Keeping: Bring Hoping Wiping Could Should Some Them Rest

Interactions Below the Surface

- Constraints Not All Created Equal
  - “The Astronomer Married the Star”
  - Selectional Restrictions Override Topic
- No Surface Regularities
  - “The emigration/immigration bill guaranteed passports to all Soviet citizens
  - No Substitute for Understanding

What is Similar

- **Ad-hoc Definitions of Sense**
  - Cluster in “word space”, WordNet Sense, “Seed Sense”: Circular
- Schutze: Vector Distance in Word Space
- Resnik: Informativeness of WordNet Subsumer + Cluster
  - Relation in Cluster not WordNet is-a hierarchy
- Yarowsky: No Similarity, Only Difference
  - Decision Lists - 1/Pair
  - Find Discriminants