Lexical Semantics & Word Sense Disambiguation

CMSC 35100

Natural Language Processing

May 15, 2003

Roadmap

- Lexical Semantics
  - Emergent Meaning
    - Thematic roles, Frames, & Primitives
- Word Sense Disambiguation
  - Selectional Restriction-based approaches
    - Limitations
  - Robust Approaches
    - Supervised Learning Approaches
      - Naïve Bayes
    - 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-internal Structure

- Thematic roles:
  - Characterize verbs by their arguments
    - E.g. transport: agent, theme, source, destination
      - They transported grain from the fields to the silo.
    - Deep structure: passive / active: same roles
- Thematic hierarchy
  - E.g. agent > theme > source, dest
    - Provide default surface positions
  - Tie to semantics (e.g. Levin): Interlinguas
    - Cluster verb meanings by set of syntactic alternations
    - Limitations: only NP,PP: other arguments predicates less

Selectional Restrictions

- Semantic constraints on filling of roles
  - E.g. Bill ate chicken
    - Eat: Agent: animate; Theme: Edible
  - Associate with sense
    - Most commonly of verb/event; possibly adj, noun…
- Specifying constraints:
  - Add a term to semantics, e.g. Isa(x,Ediblething)
  - Tie to position in WordNet
    - All hyponyms inherit

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

- 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
  - 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 Montague
- 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: 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

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?
  - 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..

Naïve Bayes' Approach

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

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

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

- "Naïve" assumption: features independent

$$P(V \mid s) = \prod_{j=1}^{n} P(v_j \mid s)$$

$$\hat{s} = \arg\max_{s \in S} P(s) \prod_{j=1}^{n} P(v_j \mid s)$$

Example: "Plant" Disambiguation

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 and many others. We use reagent injection in molten metal for the…

Industrial Example

Label the First Use of "Plant"

Yarowsky's Decision Lists: Detail

- One Sense Per Discourse - Majority
- One Sense Per Collocation
  - Near Same Words → Same Sense

Yarowsky's Decision Lists: Detail

- 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: $abs\left(\log \frac{Pr(sense_1 | Collocation)}{Pr(sense_2 | Collocation)}\right)$
    - (B) Tag New Instances with Rules
    - (C)* Apply 1 Sense/Discourse
    - (D) If Still Unlabeled, Go To 2
  - 3. Apply 1 Sense/Discouse
- Disambiguation: First Rule Matched

Sense Choice With Collocational Decision Lists

- Use Initial Decision List
  - Rules Ordered by $abs(\log \frac{Pr(sense_1|Collocation)}{Pr(sense_2|Collocation)})$
- 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

Semantic Ambiguity

- “Plant” ambiguity
  - Botanical vs Manufacturing senses
- Two types of context
  - Local: 1-2 words away
  - Global: several sentence window
- Two observations (Yarowsky 1995)
  - One sense per collocation (local)
  - One sense per discourse (global)

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: $(\mathrm{I}) = -\log(\frac{\sum_{w \in C} Count(w)}{N})$
    - Sum occurrences of it and all children
    - Informativeness
- 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 $(\mathrm{I}) = -\log(\frac{\sum_{w \in C} Count(w)}{N})$
- Result: Correct Selection

The Question of Context

- Shared Intuition:
  - Context $\longrightarrow$ 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!
- “Scratching her nose” vs “Kicking the bucket” (deMarcken 1995)
- 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

- **<u>Ad-hoc Definitions of Sense</u>**
  - 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