Coarse to Fine Grained Sense Disambiguation in Wikipedia

Hui Shen [Ohio University]
Razvan Bunescu [Ohio University]
Rada Mihalcea [University of North Texas]

*SEM, June 13, 2013
Word Senses

• A word sense is a particular meaning of a word

• Senses of a word may be entirely different with no relations, called homonyms
  – Bank: money bank, river bank

• Senses of a word may be related, called polysemes
  – Bank: financial institute, building of the financial institute, storage of blood (blood bank)

• No hard threshold to distinguish between polysemy and homonymy, it’s a matter of degree
How Many Senses a Word Has?

• Not always an easy question:
  – Drive the car
  – Drive to school
  – Drive me mad

• Can use **zeugma** to make a decision:
  – *He drives me mad and the car.*
How Many Senses a Word Has?

- Dictionaries (or humans) may differ on how many senses a word has.
- Typically dictionaries or linguistic resources give very fine-grained senses of a word, but for NLP that may not be needed (in fact that may hurt).
- WordNet has 34 senses for drive.
Relations Between Senses

• **Synonyms**: When two senses of two words are identical or very similar, e.g. buy & purchase
  – Could be tested by substitution
    • I bought/purchased a car.
  – Probably there is no perfect synonymy, they still may be different in some contexts, e.g. water and $H_2O$

• Synonymy is best defined for senses not words
  – Home purchase is a long process.
  – *Home buy is a long process.
Antonyms: Senses of words with opposite meanings, e.g. long/short, rise/fall

While antonyms are very different because they have opposite meanings, they are also very similar because they share all other aspects, e.g. long and short are degree of lengths

It is often difficult to distinguish between synonyms and antonyms if automatically extracted from a corpus using measures of context similarity

- This is good.
- This is nice.
- This is bad.
Relations Between Senses

- **Hyponyms**: A sense of a word is more specific than a sense of another word, e.g. *apple* is a hyponym of *fruit*
- **Hypernyms**: Opposite of hyponym, e.g. *fruit* is a hypernym of *apple*
- **Meronyms**: Part-whole relation, e.g. wheel is a meronym of car
- **Holonyms**: Opposite of meronyms, e.g. car is a holonym of wheel
WordNet

• A computational resource for English sense relations, lexical database

• Available for free, browse or download: http://wordnet.princeton.edu/

• Developed by famous cognitive psychologist George Miller and a team at Princeton University

• Database of word senses and their relations
WordNet

• **Synset** (synonym set): Set of near synonyms in WordNet
  – Basic primitive of WordNet
  – Each synset expresses a semantic concept
  – Example synset: {drive, thrust, driving force}

• Entry for each word shows all the synsets (senses) the word appears in, some description and sometimes example usage

• About 140,000 words and 109,000 synsets

• Synsets (not individual words) are connected by various sense relations
Some WordNet Synset Relationships

• **Antonym**: front → back
• **Similar**: unquestioning → absolute
• **Cause**: kill → die
• **Entailment**: breathe → inhale
• **Holonym**: chapter → text (part-of)
• **Meronym**: computer → cpu (whole-of)
• **Hyponym**: tree → plant (specialization)
• **Hypernym**: fruit → apple (generalization)
A WordNet Snapshot

synsets

hyponym

motor vehicle, automotive vehicle

hypernym

car, auto, automobile, machine, motorcar

meronym

accelerator, gas pedal, gas

hyponym

cab, taxi, taxicab, hack

hyponym

ambulance
WordNets for Other Languages

• EuroWordNet: Individual WordNets for some European languages (Dutch, Italian, Spanish, German, French, Czech, and Estonian) which are also interconnected by interlingual links
  http://www.illc.uva.nl/EuroWordNet/

• WordNets for some Asian languages:
  – Hindi:
    • http://www.cfilt.iitb.ac.in/wordnet/webhwn/
  – Marathi:
    • http://www.cfilt.iitb.ac.in/wordnet/webmwn/
  – Japanese:
    • http://nlpwww.nict.go.jp/wn-ja/index.en.html
WordNet Senses

• WordNets senses (like many dictionary senses) tend to be very fine-grained
• “play” as a verb has 35 senses, including
  – play a role or part: “Gielgud played Hamlet”
  – pretend to have certain qualities or state of mind: “John played dead.”
• Difficult to disambiguate to this level for people and computers. Only expert lexicographers are perhaps able to reliably differentiate senses
• Not clear such fine-grained senses are useful for NLP
• Several proposals for grouping senses into coarser, easier to identify senses (e.g. homonyms only)
Word Sense Disambiguation (WSD)

• Task of automatically selecting the correct sense for a word
• Many tasks in NLP require disambiguation of ambiguous words
  – Question Answering
  – Information Retrieval
  – Machine Translation
  – Text Mining
  – Phone Help Systems
• Understanding how people disambiguate words is an interesting problem that can provide insight in psycholinguistics
As the Earth is 4.5 billion years old, it would have lost its atmosphere by now if there were no protective magnetosphere ...

The atmosphere is composed of 78% nitrogen and 21% oxygen.

Q: What is the composition of Earth’s atmosphere?
As the Earth is 4.5 billion years old, it would have lost its atmosphere by now if there were no protective magnetosphere ...

The atmosphere is composed of 78% nitrogen and 21% oxygen.

Q: What is the composition of Earth’s atmosphere?
As the Earth is 4.5 billion years old, it would have lost its atmosphere by now if there were no protective magnetosphere ...

The atmosphere is composed of 78% nitrogen and 21% oxygen.
Word Sense Disambiguation

- Select the correct sense of a word based on the context:
  - Use a repository of senses such as WordNet:
    - Static resource, short glosses, too fine grained.
  - Unsupervised:
    - Similarity between context and sense definition or gloss.
  - Supervised:
    - Train on text manually tagged with word senses.
      - Limited amount of manually labeled data.
Unsupervised WSD using WN

Disambiguate two senses of car:

• Sense 1: auto, automobile, machine, motorcar
  Gloss: A motor vehicle with four wheels; usually propelled by an internal combustion engine.

• Sense 2: railcar, railway car, railroad car
  Gloss: A wheeled vehicle adapted to the rails of railroad.

What is the sense of car in “Three cars had jumped the rails.”?

• The sentence shares one word with the gloss of sense 2, and shares no word with the gloss of sense 1. Hence, sense 2 is chosen as the sense of car in the sentence.
Supervised WSD using WN

Disambiguate multiple senses of **fire:**
- Sense 1: the event of something burning (often destructive)
- Sense 2: the act of firing weapons or artillery at an enemy
- Sense 3: fireplace in which a relatively small fire is burning
- Sense 4: fuel that is burning and is used as a means for cooking
- Sense 5: intense adverse criticism

- They lost everything in the fire.  - Sense 1
- Hold your fire until you can see the whites of their eyes.  - Sense 2
- They sat by the fire and talked.  - Sense 3
- Put the kettle on the fire.  - Sense 4
- Clinton directed his fire at the Republican Party.  - Sense 5
Supervised WSD using WN

Features extracted from the labeled sentences are used to train multi-classes classifiers for WSD.

Some useful features:

- **Bag of Word features:**
  Neighbor words, POS of neighbor words

- **Syntactic features:**
  Syntactic relations of the target word in the parse tree:
  - The head of the target word and the relation type between the head and the target word
  - The dependents of the target word and the relation type between the target word and its dependents.
Supervised Learning for WSD

- Treat as a classification problem with the potential senses for the target word as the classification labels.
- Decide appropriate features and a classification method (Naïve Bayes, MaxEnt, decision lists etc.)
- Train using data labeled with the correct word senses
- Use the trained classifier to disambiguate instances of the target word in the test corpus
Feature Engineering

• The success of machine learning requires instances to be represented using an effective set of features that are correlated with the categories of interest
• Feature engineering can be a laborious process that requires substantial human expertise and knowledge of the domain
• In NLP it is common to extract many (even thousands of) potentially features and use a learning algorithm that works well with many relevant and irrelevant features
Contextual Features

- Surrounding bag of words
- POS of neighboring words
- Local collocations
- Syntactic relations

Experimental evaluations indicate that all of these features are useful; and the best results come from integrating all of these cues in the disambiguation process.
Surrounding Bag of Words

- Unordered individual words near the ambiguous word.
- Words in the same sentence.
- May include words in the previous sentence or surrounding paragraph.
- Gives general topical cues of the context.
- May use feature selection to determine a smaller set of words that help discriminate possible senses.
- May just remove common “stop words” such as articles, prepositions, etc.
POS of Neighboring Words

• POS of the word narrows down the senses
• Also use part-of-speech of immediately neighboring words.
• Provides evidence of local syntactic context.
• $P_{-i}$ is the POS of the word $i$ positions to the left of the target word.
• $P_i$ is the POS of the word $i$ positions to the right of the target word.
• Typical to include features for: $P_{-3}, P_{-2}, P_{-1}, P_1, P_2, P_3$
Local Collocations

• Specific lexical context immediately adjacent to the word.
• For example, to determine if “interest” as a noun refers to “readiness to give attention” or “money paid for the use of money”, the following collocations are useful:
  – “in the interest of”
  – “an interest in”
  – “interest rate”
  – “accrued interest”
• $C_{ij}$ is a feature of the sequence of words from local position $i$ to $j$ relative to the target word.
  – $C_{-2,1}$ for “in the interest of” is “in the of”
• Typical to include:
  – Single word context: $C_{-1,-1}, C_{1,1}, C_{-2,-2}, C_{2,2}$
  – Two word context: $C_{-2,-1}, C_{-1,1}, C_{1,2}$
  – Three word context: $C_{-3,-1}, C_{-2,1}, C_{-1,2}, C_{1,3}$
Syntactic Relations
(Ambiguous Verbs)

• For an ambiguous verb, it is very useful to know its direct object.
  – “played the game”
  – “played the guitar”
  – “played the risky and long-lasting card game”
  – “played the beautiful and expensive guitar”
  – “played the big brass tuba at the football game”
  – “played the game listening to the drums and the tubas”

• May also be useful to know its subject:
  – “The game was played while the band played.”
  – “The game that included a drum and a tuba was played on Friday.”
Syntactic Relations (Ambiguous Nouns)

• For an ambiguous noun, it is useful to know what verb it is an object of:
  – “played the piano and the horn”
  – “wounded by the rhinoceros’ horn”

• May also be useful to know what verb it is the subject of:
  – “the bank near the river loaned him $100”
  – “the bank is eroding and the bank has given the city the money to repair it”
Syntactic Relations (Ambiguous Adjectives)

- For an ambiguous adjective, it useful to know the noun it is modifying.
  - “a brilliant young man”
  - “a brilliant yellow light”
  - “a wooden writing desk”
  - “a wooden acting performance”
Using Syntax in WSD

- Produce a parse tree for a sentence using a syntactic parser.

```
S
  NP       VP
  ProperN V   NP
    John   played DET N
        the   piano
```

- For ambiguous verbs, use the head word of its direct object and of its subject as features.
- For ambiguous nouns, use verbs for which it is the object and the subject as features.
- For ambiguous adjectives, use the head word (noun) of its NP as a feature.
## Feature Vectors

### A small example

<table>
<thead>
<tr>
<th>Training Example</th>
<th>$P_1$</th>
<th>$P_{-1}$</th>
<th>$C_{1,1}$</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NP</td>
<td>IN</td>
<td>guitar</td>
<td>play$_1$</td>
</tr>
<tr>
<td>2</td>
<td>DT</td>
<td>RB</td>
<td>band</td>
<td>play$_2$</td>
</tr>
<tr>
<td>3</td>
<td>VBN</td>
<td>NN</td>
<td>good</td>
<td>play$_1$</td>
</tr>
<tr>
<td>4</td>
<td>DT</td>
<td>IN</td>
<td>string</td>
<td>play$_1$</td>
</tr>
</tbody>
</table>
Word Sense Disambiguation

• Select the correct sense of a word based on the context:
  – Use a repository of senses such as **WordNet**:
    • Static resource, short glosses, too fine grained.
  – **Unsupervised**:
    • Similarity between context and sense definition or gloss.
  – **Supervised**:
    • Train on text manually tagged with word senses.
      – Limited amount of manually labeled data.
Disadvantages of WSD using WN

Data sparseness:

• **Unsupervised learning:**
  The gloss for a sense is short, may share few or no word with the testing sentence, even the sentence uses the correct sense.

• **Supervised learning:**
  One sense has few sample sentences, so there are very few labeled training data for each sense of one word. This makes it difficult to train a high accurate classifier for WSD.

Too Fine grained word senses:
The distinction between different senses of one word have small degree, which make it hard for WSD.
Palermo is a city in Southern Italy, the capital of the autonomous region of Sicily.
1. Collect all WP titles that are linked from the anchor word *atmosphere*.
   => *Atmosphere, Atmosphere of Earth, Mood (Psychology), …*

2. Create a sense repository from all titles that have sufficient support in WP.
   => \{*Atmosphere, Atmosphere of Earth, Atmosphere of Mars, Atmosphere of Venus, Stellar Atmosphere, Atmosphere (unit), Atmosphere (music group)*\}

3. For each sense, use the links as labeled examples and train a classifier to distinguish between alternative senses of the word *atmosphere*:
   – extract features from the word context.
   – each WP sense acts as a different label in the classification model.
Major Shortcoming of WP Corpus for WSD

• Classification algorithms work with **disjoint categories**.

• **Sense labels collected from WP are not disjoint!**
  
  – Many instances that are linked to *Atmosphere* could have been linked to more specific titles: *Atmosphere of Earth* or *Atmosphere of Mars*.

  • *Atmosphere* category is ill defined.

  ⇒ the learning algorithm will underperform, since it tries to separate *Atmosphere* examples from *Atmosphere of Earth* examples.
The Beagle 2 lander objectives were to characterize the physical properties of the atmosphere and surface layers.
The Orbiter has been successfully performing scientific measurements since early 2004, namely high-resolution imaging and study of the interaction of the atmosphere with the interplanetary medium.

Atmosphere of Mars

Atmosphere

reference

sense
Assuming the planet’s atmosphere is close to chemical equilibrium, it is predicted that 55 Cancri d is covered in a layer of water clouds.
An aerogravity assist, or AGA, is a spacecraft maneuver designed to change velocity when arriving at a body with an atmosphere.
Annotation Inconsistencies in Wikipedia: Potential Causes

1. Editors may be unaware that an article exists in Wikipedia for the actual reference of a word, or for a more specific sense of the word:
   - end up using a link to an article describing the general sense of the word.

2. More specific articles are introduced only in newer versions of Wikipedia:
   - and thus earlier annotations could not have been aware of these more recent articles.
3. Annotating words with the most specific sense or reference available in WP may require substantial cognitive effort:
   – editors will then choose to link to a general sense of the word.

Trajan was nominated as Consul and brought Apollodorus of Damascus with him to Rome around 91.
An animal sleeps on the couch.

A cat sleeps on the couch.

A white siameze cat sleeps on the couch.
A cat sleeps on the piece of furniture.

A cat sleeps on the couch.

A cat sleeps on the white leather couch.
Possible Solutions (1)

1) Group overlapping senses and references into one general sense category.

=> {Atmosphere, Atmosphere (unit), Atmosphere (music group)}
Possible Solutions (1)

• **Straightforward to implement:**
  – Train multiclass classifier to distinguish the disjoint categories 
    \{Atmosphere, Atmosphere (unit), Atmosphere (music group)\}

• **Disadvantage is loss of disambiguation resolution:**
  – Resulting WSD system cannot link atmosphere to the more specific senses 
    \{Atmosphere of Earth, Atmosphere of Mars, Atmosphere of Venus, Stellar Atmosphere\}
Possible Solutions (2)

2) Keep the original sense repository, but change the definition of some sense categories such that all categories in the repository become mutually disjoint.
Possible Solutions (2)

2) Keep the original sense repository, but change the definition of some sense categories such that all categories in the repository become mutually disjoint.
Atmosphere (G) = *generic uses, or senses/references not in WP.*

=> \{Atmosphere (G), Atmosphere of Earth, Atmosphere of Mars, Atmosphere of Venus, Stellar Atmosphere, Atmosphere (unit), Atmosphere (music group)\}
Possible Solutions (2)

• Advantage is WSD system can make more fine grained annotations:
  – annotate senses down to reference level i.e. Atmosphere of Earth, Atmosphere of Mars, Stellar Atmosphere.

• Not as straightforward to train:

\[
\text{Atmosphere (G)}
\]

known set of training examples  \hspace{2cm}  unknown set of training examples
“In global climate models, the properties of the atmosphere are specified at a number of …”
“In global climate models, the properties of the atmosphere are specified at a number of …”
“In global climate models, the properties of the **atmosphere** are specified at a number of …”
“In global climate models, the properties of the atmosphere are specified at a number of …”
“In global climate models, the properties of the atmosphere are specified at a number of …”

Proposed WSD Architecture

Level 1
Atmosphere (unit) — Atmosphere — Atmosphere (music group)

Level 2
Atmosphere (G) — Atmosphere (S)

Level 3
A. Mars — A. Earth — A. Venus — Stellar A.
Proposed WSD Architecture

“In global climate models, the properties of the atmosphere are specified at a number of …”
Straightforward Training WSD for L1 & L3: Use multiclass classifiers

Level 1
Atmosphere (unit) ➔ Atmosphere ➔ Atmosphere (music group)

training examples
Atmosphere, Atmosphere of Earth, Atmosphere of Mars, Atmosphere of Venus, Stellar Atmosphere

Level 3
A. Mars ➔ A. Earth ➔ A. Venus ➔ Stellar A.
Training Level 2 Classifier

Level 2

Atmosphere (G)  Atmosphere (S)

training examples

Atmosphere of Earth, Atmosphere of Mars, Atmosphere of Venus, Stellar Atmosphere

known set of training examples

unknown set of training examples

?
Naïve SVM

- 60% of the examples linked to Atmosphere should actually belong to Atmosphere (S).
  ⇒ underperforming classifier.
Training for Level 2: Semi-supervised

unlabeled training examples
Atmosphere

positive training examples
Atmosphere of Earth, Atmosphere of Mars,
Atmosphere of Venus, Stellar Atmosphere

⇒ learning with positive and unlabeled examples.
Learning with Standard SVM

1) Standard SVM:

- class labels
- discriminant function
- slack variables
- upper bound of number of miss-classified data
Learning with Positive and Unlabeled Examples

1) Adaptation of the Biased SVM [Lee and Liu, ICML 2003]:

\[
\text{minimize: } \frac{1}{2} \|w\|^2 + C_P \sum_{x \in P} \xi_x + C_U \sum_{x \in U} \xi_x \\
\text{subject to: } s(x) (w^T \phi(x) + b) \geq 1 - \xi_x, \, \forall x \in P \cup U \\
\xi_x \geq 0
\]

\[s(x) = +1, \text{ if } x \text{ positive (P)}\]
\[s(x) = -1, \text{ if } x \text{ unlabeled (U)}\]
Learning with Positive and Unlabeled Examples

1) Adaptation of the Biased SVM [Lee and Liu, ICML 2003]:

\[
\text{minimize: } \frac{1}{2} \|w\|^2 + C_P \sum_{x \in P} \xi_x + C_U \sum_{x \in U} \xi_x \\
\text{subject to: } s(x) (w^T \phi(x) + b) \geq 1 - \xi_x, \quad \forall x \in P \cup U \\
\xi_x \geq 0
\]

- control penalties for errors on positive ($C_P$) vs. unlabeled ($C_U$) examples
- how do we set $C_P$ and $C_U$?
  - want $C_P > C_U$
1) Adaptation of Biased SVM [Lee and Liu, ICML 2003]:

\[
\text{minimize: } \frac{1}{2} \|w\|^2 + C_P \sum_{x \in P} \xi_x + C_U \sum_{x \in U} \xi_x \\
\text{subject to: } s(x)(w^T \phi(x) + b) \geq 1 - \xi_x, \quad \forall x \in P \cup U \\
\xi_x \geq 0
\]

- Use development data to tune $C_P$ and $C_U$.
- $\arg\max pr = \arg\max_{C_P,C_U} r^2 / p(f = 1)$ [Lee and Liu, 03]
- $\arg\max acc = \arg\max_{C_P,C_U} \text{recall} - p(f = 1)$ [this work, 13]

both estimated on positive and unlabeled examples from dev. data
Learning with Positive and Unlabeled Examples

2) Weighted Samples SVM [Elkan and Noto, KDD 2008]:

I. Train prob. classifier $g(x)$ to compute $p(s = 1|x)$ on $P$ and $U$.

II. Train final decision function $f(x)$ on weighted examples sampled as follows:

- sample each example in $P$ with weight 1.
- sample each example from $U$ as two examples:
  - one positive, with weight $p(y = +1 | x, s = 0)$.
  - one negative, with weight $p(y = -1 | x, s = 0)$.

both estimated based on $g(x)$
# Evaluation Datasets

<table>
<thead>
<tr>
<th>Atmosphere</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmosphere of Earth</td>
<td>518</td>
</tr>
<tr>
<td>Atmosphere of Mars</td>
<td>19</td>
</tr>
<tr>
<td>Atmosphere of Venus</td>
<td>9</td>
</tr>
<tr>
<td>Stellar Atmosphere</td>
<td>13</td>
</tr>
<tr>
<td>Atmosphere (O)</td>
<td>373</td>
</tr>
<tr>
<td>Atmosphere of Earth (S)</td>
<td>932</td>
</tr>
<tr>
<td>Atmosphere of Mars (S)</td>
<td>559</td>
</tr>
<tr>
<td>Atmosphere of Venus (S)</td>
<td>11</td>
</tr>
<tr>
<td>Stellar Atmosphere (S)</td>
<td>29</td>
</tr>
<tr>
<td>Atmosphere (UNIT)</td>
<td>96</td>
</tr>
<tr>
<td>Atmosphere (MUSIC GROUP)</td>
<td>104</td>
</tr>
</tbody>
</table>
## Evaluation Datasets

<table>
<thead>
<tr>
<th><strong>president</strong></th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>President</strong></td>
<td>3534</td>
</tr>
<tr>
<td><em>President (S)</em></td>
<td>989</td>
</tr>
<tr>
<td>Chancellor (education)</td>
<td>326</td>
</tr>
<tr>
<td>President of the United States</td>
<td>534</td>
</tr>
<tr>
<td>President of the Philippines</td>
<td>42</td>
</tr>
<tr>
<td>President of Pakistan</td>
<td>27</td>
</tr>
<tr>
<td>President of France</td>
<td>22</td>
</tr>
<tr>
<td>President of India</td>
<td>21</td>
</tr>
<tr>
<td>President of Russia</td>
<td>17</td>
</tr>
<tr>
<td><strong>President (O)</strong></td>
<td>2545</td>
</tr>
<tr>
<td>Chancellor (education)</td>
<td>210</td>
</tr>
<tr>
<td>President of the United States</td>
<td>5941</td>
</tr>
<tr>
<td>President of the Philippines</td>
<td>549</td>
</tr>
<tr>
<td>President of Pakistan</td>
<td>192</td>
</tr>
<tr>
<td>President of France</td>
<td>151</td>
</tr>
<tr>
<td>President of India</td>
<td>86</td>
</tr>
<tr>
<td>President of Russia</td>
<td>101</td>
</tr>
</tbody>
</table>
## Evaluation Datasets

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dollar</strong></td>
<td></td>
</tr>
<tr>
<td>Dollar (S)</td>
<td>231</td>
</tr>
<tr>
<td>United States dollar</td>
<td>228</td>
</tr>
<tr>
<td>Canadian dollar</td>
<td>3</td>
</tr>
<tr>
<td>Australian dollar</td>
<td>1</td>
</tr>
<tr>
<td>Dollar (O)</td>
<td>147</td>
</tr>
<tr>
<td><strong>United States dollar</strong></td>
<td>3516</td>
</tr>
<tr>
<td><strong>Canadian dollar</strong></td>
<td>420</td>
</tr>
<tr>
<td><strong>Australian dollar</strong></td>
<td>124</td>
</tr>
<tr>
<td><strong>Dollar sign</strong></td>
<td>290</td>
</tr>
<tr>
<td><strong>Dollar (band)</strong></td>
<td>30</td>
</tr>
<tr>
<td><strong>Dollar, ClackMannanshire</strong></td>
<td>30</td>
</tr>
<tr>
<td><strong>Game</strong></td>
<td></td>
</tr>
<tr>
<td>Game (S)</td>
<td>819</td>
</tr>
<tr>
<td>Video game</td>
<td>99</td>
</tr>
<tr>
<td>PC game</td>
<td>55</td>
</tr>
<tr>
<td>Game (O)</td>
<td>720</td>
</tr>
<tr>
<td><strong>Video game</strong></td>
<td>312</td>
</tr>
<tr>
<td><strong>PC game</strong></td>
<td>24</td>
</tr>
<tr>
<td><strong>Game (food)</strong></td>
<td>232</td>
</tr>
<tr>
<td><strong>Game (rapper)</strong></td>
<td>154</td>
</tr>
</tbody>
</table>
## Evaluation Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diamond</strong></td>
<td></td>
</tr>
<tr>
<td>Diamond</td>
<td>716</td>
</tr>
<tr>
<td>Diamond (S)</td>
<td>221</td>
</tr>
<tr>
<td>Diamond (gemstone)</td>
<td>221</td>
</tr>
<tr>
<td>Diamond (G)</td>
<td>495</td>
</tr>
<tr>
<td><strong>Diamond (Gemstone)</strong></td>
<td>71</td>
</tr>
<tr>
<td><strong>Baseball Field</strong></td>
<td>36</td>
</tr>
<tr>
<td><strong>Music Recording Sales Cert.</strong></td>
<td>36</td>
</tr>
<tr>
<td><strong>Corinth</strong></td>
<td></td>
</tr>
<tr>
<td>Corinth</td>
<td>699</td>
</tr>
<tr>
<td>Corinth (S)</td>
<td>409</td>
</tr>
<tr>
<td>Ancient Corinth</td>
<td>409</td>
</tr>
<tr>
<td>Corinth (G)</td>
<td>290</td>
</tr>
<tr>
<td><strong>Ancient Corinth</strong></td>
<td>92</td>
</tr>
<tr>
<td><strong>Corinth, Mississippi</strong></td>
<td>72</td>
</tr>
</tbody>
</table>
### Experimental Results: Level 2 Accuracy

<table>
<thead>
<tr>
<th>Word</th>
<th>NaiveSVM</th>
<th>BiasedSVM</th>
<th>WeightedSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>atmosphere</td>
<td>39.9%</td>
<td>79.6%</td>
<td>75.0%</td>
</tr>
<tr>
<td>president</td>
<td>91.9%</td>
<td>92.5%</td>
<td>89.5%</td>
</tr>
<tr>
<td>dollar</td>
<td>96.0%</td>
<td>97.0%</td>
<td>97.1%</td>
</tr>
<tr>
<td>game</td>
<td>83.8%</td>
<td>87.1%</td>
<td>84.6%</td>
</tr>
<tr>
<td>diamond</td>
<td>70.2%</td>
<td>74.5%</td>
<td>75.1%</td>
</tr>
<tr>
<td>Corinth</td>
<td>46.2%</td>
<td>75.1%</td>
<td>51.9%</td>
</tr>
<tr>
<td>presidentₜ</td>
<td>88.1%</td>
<td>90.6%</td>
<td>87.4%</td>
</tr>
<tr>
<td>dollarₜ</td>
<td>70.3%</td>
<td>84.9%</td>
<td>70.6%</td>
</tr>
</tbody>
</table>

- Trained only on WP links, tested on manual annotations.
  - Averaged over 4-fold cross-validation experiments.
**Experimental Results: Level 2 F-measure**

<table>
<thead>
<tr>
<th>Word</th>
<th>NaiveSVM</th>
<th>BiasedSVM</th>
<th>WeightedSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>atmosphere</td>
<td>30.5%</td>
<td>86.0%</td>
<td>83.2%</td>
</tr>
<tr>
<td>president</td>
<td>94.4%</td>
<td>95.0%</td>
<td>92.8%</td>
</tr>
<tr>
<td>dollar</td>
<td>97.9%</td>
<td>98.4%</td>
<td>98.5%</td>
</tr>
<tr>
<td>game</td>
<td>75.1%</td>
<td>81.8%</td>
<td>77.5%</td>
</tr>
<tr>
<td>diamond</td>
<td>8.6%</td>
<td>53.5%</td>
<td>46.3%</td>
</tr>
<tr>
<td>Corinth</td>
<td>15.3%</td>
<td>81.2%</td>
<td>68.0%</td>
</tr>
</tbody>
</table>
| president,
             | 90.0%    | 92.4%     | 89.5%       |
| dollar,
             | 77.9%    | 91.2%     | 78.2%       |

- Trained only on WP links, tested on manual annotations.
  - Averaged over 4-fold cross-validation experiments.
Experimental Results: Overall Accuracy

...atmosphere...

Level 1
- Atmosphere (unit)
- Atmosphere
- Atmosphere (music group)

Level 2
- Atmosphere (G)
- Atmosphere (S)

Level 3
- A. Mars
- A. Earth
- A. Venus
- Stellar A.
Experimental Results: Overall Accuracy

<table>
<thead>
<tr>
<th></th>
<th>atmosphere</th>
<th>president</th>
<th>dollar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat Hierarchical</td>
<td>52.4%</td>
<td>89.4%</td>
<td>90.0%</td>
</tr>
<tr>
<td>Flat Hierarchical</td>
<td>79.7%</td>
<td>91.0%</td>
<td>90.1%</td>
</tr>
<tr>
<td>Flat Hierarchical</td>
<td>game</td>
<td>diamond</td>
<td>Corinth</td>
</tr>
<tr>
<td>Flat Hierarchical</td>
<td>83.6%</td>
<td>65.7%</td>
<td>42.6%</td>
</tr>
<tr>
<td>Flat Hierarchical</td>
<td>87.2%</td>
<td>76.8%</td>
<td>72.1%</td>
</tr>
</tbody>
</table>

- Leaf nodes as sense repository.
Future Work

• Extend learning with positive and unlabeled data to multiclass:

  “In global climate models, the properties of the **atmosphere** are specified at a number of …”

• Automatically add fine grained links to entire Wikipedia.
Questions