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
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.

- Use Wikipedia for WSD:
  - Large sense repository, continuously growing.
  - Large training dataset.
"Palermo" is a city in [Southern Italy], the [capital city | capital] of the [autonomous area | autonomous region] of Sicily.

**Palermo** is a city in **Southern Italy**, the **capital** of the **autonomous region** of **Sicily**.

capital (economics)  capital city  human capital

capital (architecture)  financial capital
WSD Trained on Wikipedia Corpus
– atmosphere –

1. Collect all WP titles that are linked from the anchor word atmosphere.
   => Atmospheric, 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.
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.
Annotation Inconsistencies in Wikipedia: Potential Causes

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.
An 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\}
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:

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

\text{known set of training examples} \quad \text{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 \textit{atmosphere} are specified at a number of …”
“In global climate models, the properties of the **atmosphere** are specified at a number of ...”
Straightforward Training WSD for L1 & L3: Use multiclass classifiers

training examples
Atmosphere, Atmosphere of Earth, Atmosphere of Mars, Atmosphere of Venus, Stellar Atmosphere
Training Level 2 Classifier

Atmosphere (G)

Atmosphere (S)

training examples

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

unknown set of training examples

known set of training examples
Naïve SVM

Level 2

Atmosphere (G) ➔ Atmosphere (S)

training examples

Atmosphere

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

• 60% of the examples linked to Atmosphere should actually belong to Atmosphere (S).

⇒ underperforming classifier.
Training for Level 2: Semi-supervised

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

⇒ learning with positive and unlabeled examples.
1) Adaptation of the Biased SVM [Lee and Liu, ICML 2003]:

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

\[
s(x) = +1, \text{ if } x \text{ positive (P)} \\
s(x) = -1, \text{ if } x \text{ unlabeled (U)}
\]

decision function \( f(x) \)
1) Adaptation of the Biased SVM [Lee and Liu, ICML 2003]:

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

- 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]:

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

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

*both estimated un positive and unlabeled examples from dev. data*
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

| atmosphere                      | Size |
|---------------------------------|------|
| **ATMOSPHERE**                  |      |
| Atmosphere (S)                  |      |
| Atmosphere of Earth             | 518  |
| Atmosphere of Mars              | 19   |
| Atmosphere of Venus             | 9    |
| Stellar Atmosphere              | 13   |
| Atmosphere (O)                  | 373  |
| **ATMOSPHERE OF EARTH**         |      |
| **ATMOSPHERE OF MARS**          |      |
| **ATMOSPHERE OF VENUS**         |      |
| **STELLAR ATMOSPHERE**          |      |
| **ATMOSPHERE (UNIT)**           |      |
| **ATMOSPHERE (MUSIC GROUP)**    |      |
|                                 | 345  |
|                                 | 37   |
|                                 | 26   |
|                                 | 29   |
|                                 | 96   |
|                                 | 104  |
## Evaluation Datasets

| president                                      | Size  |
|------------------------------------------------|-------|
| **President**                                   | 3534  |
| *President (S)*                                 | 989   |
| Chancellor (education)                         | 326   |
| President of the United States                 | 534   |
| President of the Philippines                   | 42    |
| President of Pakistan                          | 27    |
| President of France                            | 22    |
| President of India                             | 21    |
| President of Russia                            | 17    |
| **President (O)**                              | 2545  |
| **Chancellor (education)**                      | 210   |
| President of the United States                 | 5941  |
| President of the Philippines                   | 549   |
| President of Pakistan                          | 192   |
| President of France                            | 151   |
| President of India                             | 86    |
| President of Russia                            | 101   |
# Evaluation Datasets

| dollar                  | Size |
|-------------------------|------|
| **DOLLAR**              | 379  |
| Dollar (S)              | 231  |
| United States dollar    | 228  |
| Canadian dollar         | 3    |
| Australian dollar       | 1    |
| Dollar (O)              | 147  |
| **UNITED STATES DOLLAR**| 3516 |
| **CANADIAN DOLLAR**     | 420  |
| **AUSTRALIAN DOLLAR**   | 124  |
| **DOLLAR SIGN**         | 290  |
| **DOLLAR (BAND)**       | 30   |
| **DOLLAR, CLACKMANNANSHERE** | 30 |

| game                    | Size |
|-------------------------|------|
| **GAME**                | 819  |
| Game (S)                | 99   |
| Video game              | 55   |
| PC game                 | 44   |
| Game (O)                | 720  |
| **VIDEO GAME**          | 312  |
| **PC GAME**             | 24   |
| **GAME (FOOD)**         | 232  |
| **GAME (RAPPER)**       | 154  |
## Evaluation Datasets

| dataset                          | size  |
|---------------------------------|-------|
| **diamond**                     |       |
| **Diamond**                     | 716   |
| *Diamond (S)*                   | 221   |
| *Diamond (gemstone)*            | 221   |
| *Diamond (G)*                   | 495   |
| **Diamond (gemstone)**          | 71    |
| **Baseball field**              | 36    |
| **Music recording sales cert.** | 36    |
| **Corinth**                     |       |
| **Corinth**                     | 699   |
| *Corinth (S)*                   | 409   |
| *Ancient Corinth*               | 409   |
| *Corinth (G)*                   | 290   |
| **Ancient Corinth**             | 92    |
| **Corinth, Mississippi**        | 72    |
## Experimental Results: Level 2 Accuracy

### Table

| Word     | NaiveSVM | BiasedSVM | WeightedSVM |
|----------|----------|-----------|-------------|
| atmosphere | 39.9%    | 79.6%     | 75.0%       |
| president | 91.9%    | 92.5%     | 89.5%       |
| dollar    | 96.0%    | 97.0%     | 97.1%       |
| game      | 83.8%    | 87.1%     | 84.6%       |
| diamond   | 70.2%    | 74.5%     | 75.1%       |
| Corinth   | 46.2%    | 75.1%     | 51.9%       |
| president$_S$ | 88.1% | 90.6%     | 87.4%       |
| dollar$_S$ | 70.3%    | 84.9%     | 70.6%       |

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

| Word     | NaiveSVM | BiasedSVM | WeightedSVM |
|----------|----------|-----------|-------------|
| atmosphere | 30.5%    | 86.0%     | 83.2%       |
| president | 94.4%    | 95.0%     | 92.8%       |
| dollar    | 97.9%    | 98.4%     | 98.5%       |
| game      | 75.1%    | 81.8%     | 77.5%       |
| diamond   | 8.6%     | 53.5%     | 46.3%       |
| Corinth   | 15.3%    | 81.2%     | 68.0%       |
| president$_S$ | 90.0% | 92.4%     | 89.5%       |
| dollar$_S$ | 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

|                     | atmosphere | president | dollar  |
|---------------------|------------|-----------|---------|
| Flat                | 52.4%      | 89.4%     | 90.0%   |
| Hierarchical        | 79.7%      | 91.0%     | 90.1%   |
| game                |            | diamond   | Corinth |
| Flat                | 83.6%      | 65.7%     | 42.6%   |
| Hierarchical        | 87.2%      | 76.8%     | 72.1%   |

- 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