Word Sense Disambiguation with Pictures

Presented by: Daniel Crockett and Xiaoxia Wu
Word Sense Ambiguity

Some Examples:

“27 dental hygiene students to receive caps at MCC”

“Iraqi Head Seeks Arms”

“Drunk Gets Nine Months in Violin Case”
Word Senses

The noun “bass” has 8 senses in WordNet.
1. bass¹ - (the lowest part of the musical range)
2. bass², bass part¹ - (the lowest part in polyphonic music)
3. bass³, basso¹ - (an adult male singer with the lowest voice)
4. sea bass¹, bass⁴ - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass¹, bass⁵ - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass⁶, bass voice¹, basso² - (the lowest adult male singing voice)
7. bass⁷ - (the member with the lowest range of a family of musical instruments)
8. bass⁸ - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective “bass” has 1 sense in WordNet.
1. bass¹, deep⁶ - (having or denoting a low vocal or instrumental range)
   "a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"
Word Sense Disambiguation (WSD)

- Done by examining surrounding words within a window of n (bag-of-words)
- May include information about position of surrounding words (collocation features)

  “I like going to the river to catch bass”
  “I opened an account at the bank.”
Idea of paper is to use images for word sense disambiguation

- Human’s use images to resolve ambiguity
- Sense may be very clear from image, ambiguous from text
- Text with images are common
Images are frequently attached to text:

**US & Canada**

**Colorado runner kills cougar in self-defence after attack**

1 hour ago

Three things to look for in Trump's speech

7 hours ago

Boy called Trump invited to State of Union

8 hours ago

Features

Three things to look for in Trump's speech

Mountain lion attacks in North America are very rare. The image

A man running on a popular park trail in the mountains of northern Colorado killed a mountain lion after it pounced on him from behind.

Colorado Parks and Wildlife (CPW) officials say the man sustained serious injuries after he was bitten on his face and wrist by the young male lion.

The man, who has not been named, turned after hearing a noise behind him, just as the lion lunged, officials say.
Methods of Disambiguation

1. Supervised
   (Decision Trees, Naives Bayes, Neural Networks, Support Vector Machines, etc.)

2. Unsupervised
3. Knowledge-base
Methods of Disambiguation: Bayes Methods

Naives Bayes (Statistical NLP Methods)

-- Textual Content (a semantic problem, Brown et al, 1991)

-- Images (Barnard & Johnson, 2005)
Methods of Disambiguation: Bayes with Images

- **Assumption 1**: The words are linked to images via regions
- **Assumption 2**: Image and associated text are generated from multiple draws from the hidden factors

\[
P(w, b) = \sum_l P(w \mid l) P(b \mid l) P(l)
\]

- a particular blob (images):
  - a region, with its feature vector (color, texture, size, position, etc)
- indexes over the concepts
- a word
- a frequency table
- a Gaussian distribution over the features with **Diagonal Covariance Matrix**
Methods of Disambiguation: Bayes with Images

\[ P(w, b) = \sum_l P(w | l) P(b | l) P(l) \]

- A blob oriented expression \( \rightarrow \) the entire image

**Conditional Independence Assumption!**

\[ P(W \cup B) = \prod_{b \in B} \left( \sum_l P(b | l) P(l) \right)^{\max(N_b)} 
\prod_{w \in W} \left( \sum_l P(w | l) P(l | B) \right)^{\max(N_w)} \]

- Expectation Maximization (parameters for the conditional probabilities)
Methods of Disambiguation: Constrained Word Prediction
The Novel Algorithm to WSD

- Predict word from image only

but how about with associated words?

- Just use the KNOWN vocabulary
  - Automatic Region Label images
  - Reduce the ambiguity
Methods of Disambiguation: Constrained Word Prediction Link to WSD

- Just use the KNOWN vocabulary: **Reduce the ambiguity**

- A set of the *disambiguated vocabulary* \( S \),

  \[
  \text{bank} \in W \text{ then } \{ \text{bank}_1, \text{bank}_2, \ldots \} \in S.
  \]

  Assume the image provides information independent of the text method

  \[
  P(s|w, B, W) \propto P(s|w, B)P(s|w, W).
  \]
Iterative Word Sense Disambiguation

The algorithm:
1. Create $SDW = \{\}$ and $SAW = \{\text{(all words)}\}$

2. Move proper nouns to $SDW$

3. Move monosemous words to $SDW$ with their sense

4. If probability of sense based on neighbors greater than other probabilities of other senses by threshold, move to $SDW$ with sense

5. Create “noun contexts” for all noun senses in data set. Move nouns in $SAW$ that have overlap with a sense’s noun context that is greater than other sense’s overlap by a certain threshold to $SDW$
Iterative Word Sense Disambiguation

6. For every word in SAW, if semantic distance (based on WordNet) is zero from word in SDW, move to SDW with that sense

7. Repeat above, but use words in SAW (two SAW words with distance 0 in WordNet are given that WordNet sense)

8. Repeat 6 and 7 with a distance of 1

9. For any words still in SAW, assign the most common word sense to them
Iterative Word Sense Disambiguation

- Outdated
- Methods using neural networks are currently achieving state-of-the-art
- “Semi-supervised Word Sense Disambiguation with Neural Models” and “Breaking Sticks and Ambiguities with Adaptive Skip-Gram” are two more recent papers dealing with WSD.
Combining the Predictions

The predictions are assumed to be independent

\[ P(s|w, B, W) \propto P(s|w, B)P(s|w, W) \]

This is likely not true, but empirically works well.
Creating the Corpus
Experimentation

- 90/10 split on training/testing
- Run 20 times, with different splits
- Word sense occurring less than 20 (or 50) times removed
- Performance based on documents with at least one ambiguous word


Results

Combining word sense predictions from images with traditional method improves the accuracy

<table>
<thead>
<tr>
<th>Data set</th>
<th>Minimum sense count</th>
<th>Baseline</th>
<th>Text only using [41]</th>
<th>Image only</th>
<th>Combined (using (6))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>20</td>
<td>0.615</td>
<td>0.683</td>
<td>0.791</td>
<td>0.817</td>
</tr>
<tr>
<td>Full</td>
<td>50</td>
<td>0.606</td>
<td>0.674</td>
<td>0.781</td>
<td>0.814</td>
</tr>
<tr>
<td>Seed</td>
<td>20</td>
<td>0.571</td>
<td>0.693</td>
<td>0.687</td>
<td>0.741</td>
</tr>
</tbody>
</table>
Example of Model Working

Adding image data corrected senses of “plant” and “water”

Sense tagged words around plant:
rooting_1  developed_1  compost_1  sand_1
benefit_1  good_1  find_1  day_3  feel_2  separate_1  top_2  half_1  plant_2

Sense tagged words around water:
reach_1  location_1  sundown_1  herd_1
water_2  and_then_1  broad_1  grass_1  flat_1
Issues with the Paper

- Algorithms are outdated
- Possible that with better algorithms for WSD contribution from images will be negligible
- Corpus is somewhat artificial
Contributions of Paper

- ImCor
- Demonstrated potential of resolving ambiguity with images
- Demonstrated combining predictions as independent works well
Research Possibilities

- Replicate experiment with current models for images and WSD
- Incorporating image data for machine translation of websites
- Test on dataset created from real-life data
Questions?