Identifying Negative Exemplars in Grounded Language Data Sets

Nisha Pillai, Cynthia Matuszek
{ npillai1 | cmat } @ umbc.edu
Task

- Learn the meaning of a word from natural conversation despite not having negative examples
  - Learn an association from language to perceived environment
  - Visual percepts $\leftrightarrow$ attribute words
  - Joint model of visual percepts and natural language to identify novel object, shape, and color described by tokens (words)\(^{[1]}\)

- Obtain important positive terms to learn

- Find appropriate negative examples
  - statistical language comparison metrics

\[\text{Blue (positive example)} \quad \text{Not blue (negative example)}\]

\(^{[1]}\) Matuszek, FitzGerald, Zettlemoyer, Bo, Fox. ICML 2012.
Motivation

- Unavailability of negative examples in natural conversation
  
  “this is a lemon”
  “this object is an yellow ball”
  “this is not a carrot”

- Difficulty in gathering negative information without prompting

- Lack of positive label may not be a negative!
  
  “this is a lemon” $\Rightarrow$ “not yellow”
Goals

- Choose words to learn
  - Relevant, semantically meaningful, important
- Find an efficient way of obtaining negative examples
- Measure effectiveness of choices for language acquisition

Choose to train “banana” classifier

Positive example

Negative example
Grounding Training

- Training visual classifiers based on percepts
- When new language tokens are encountered:
  - Important tokens selected
  - Visual classifiers created and trained on perceptual context
- As more objects are seen, ‘best’ classifier emerge
  - E.g., most predictive of data observed so far

Language Annotation
“This is a short green cube.”

Perceived world state

Newly created semantics

Word “cube”

NEW-CLASSIFIER-CALLED-‘cube’
Data Corpus Collection

- 72 classes, 18 categories
  - Food objects
  - Children’s blocks

- Descriptive Language:
  - 3055 descriptions from Mechanical Turk
  - 19,947 unique words
    - 200-450 words/document

- 230 unique tokens selected for learning
Approach Overview

- **Dataset**: real world objects (toys, food)
- **Language**: crowdsourced human descriptions
- **Documents**: set of a descriptions of each object
- **Positive labels**: visually meaningful words worth learning
- **Negative examples**: objects chosen as negatives for them
**Approach Overview 2**

**Objects**
- "It is a lime"
- "A green ball"
- "this is a red cube"
- "a square cuboid"
- "a green arch"
- "an arch-shaped thing"

**Descriptions**
- d1: \( t_1 = \text{"lime"}, t_2 = \text{"green"}, t_3 = \text{"ball"} \)
- d2: \( t_1 = \text{"red"}, t_2 = \text{"cube"}, t_3 = \text{"square"}, t_4 = \text{"cuboid"} \)
- d3: \( t_1 = \text{"green"}, t_2 = \text{"arch"}, t_3 = \text{"arch-shaped"}, t_4 = \text{"thing"} \)

**Descriptive documents, D**

**TF-IDF**
- \( t_1 d_1 = 11 \)
- \( t_2 d_1 = 13 \)
- \( t_3 d_1 = 7 \)
- \( t_4 d_3 = 4 \)

**Positive Terms**
- "lime" "green"
- "red" "cube"
- "green" "arch"

**Cosine Similarity**
- SIM(PV(d1), PV(d2))
- SIM(PV(d1), PV(d3))

**Visual classifier denoted by "Green"**

**Threshold**
Choosing Words to Learn

- Positive labels: choosing visually meaningful words to train classifiers for

- tf-idf: term frequency-inverse document frequency
  - How important a word is to a document
  - Increases proportionally to the number of times a term appears in the document
  - Decreases with the number of documents containing that term

\[
tf-idf(t, d, D) = tf(t, d) \cdot \log \frac{N}{|\{d \in D : t \in d\}|}
\]

- \(tf(t, d)\) - the number of times a term \(t\) appears in document, \(d\).
- \(|\{d \in D : t \in d\}|\) - the number of documents in which the term \(t\) appears.
- \(N\) - the size of the set of documents. \(|D|\)
**Document Features**

**Negative examples:** semantically distant objects using Paragraph Vector\[^{[2]}\] and cosine distance

- Log probability vector
  \[ y = b + Uh(w_{t-k}, \ldots, w_{t+k}; W, D) \]
  - \(U, b\) – Softmax parameters
  - \(h\) – average of \(W\)'s and \(D\)
  - \(k\) – context window parameter

Learning using softmax classifier

\[
p(w_t | w_{t-k}, \ldots, w_{t+k}) = \frac{e^{y_{w_t}}}{\sum_i e^{y_i}}
\]

Maximize average log probability:

\[
\frac{1}{T} \sum_{t=k}^{T-k} \log p(w_t | w_{t-k}, \ldots, w_{t+k})
\]

- Cosine similarity:

\[
\cos(\theta) = \frac{A \cdot B}{\|A\|_2 \|B\|_2}
\]

- cosine of angle between documents in vector space

\[^{[2]}\] Quoc Le and Tomas Mikolov. ICML 2014.
Term Selection

- 57 top terms
- Human errors
  - Tomato / tomatoe
  - Eggplant / eggplanet

- tf-idf positive / negative labels
  - “Arch” is negative 😞
  - PV-DM fixes this
Choosing Negatives

- Vectors $\rightarrow$ individual objects
- Angle $\rightarrow$ similarity of descriptions
## Example Results

<table>
<thead>
<tr>
<th>Label</th>
<th>Positive Examples</th>
<th>Negative Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>“carrot”</td>
<td><img src="carrot1.png" alt="Carrot Examples" /> <img src="carrot2.png" alt="Carrot Examples" /> <img src="carrot3.png" alt="Carrot Examples" /></td>
<td><img src="negative1.png" alt="Negative Examples" /> <img src="negative2.png" alt="Negative Examples" /> <img src="negative3.png" alt="Negative Examples" /></td>
</tr>
<tr>
<td>“rectangular”</td>
<td><img src="rectangular1.png" alt="Rectangular Examples" /> <img src="rectangular2.png" alt="Rectangular Examples" /> <img src="rectangular3.png" alt="Rectangular Examples" /></td>
<td><img src="negative1.png" alt="Negative Examples" /> <img src="negative2.png" alt="Negative Examples" /> <img src="negative3.png" alt="Negative Examples" /></td>
</tr>
<tr>
<td>“red”</td>
<td><img src="red1.png" alt="Red Examples" /> <img src="red2.png" alt="Red Examples" /> <img src="red3.png" alt="Red Examples" /></td>
<td><img src="negative1.png" alt="Negative Examples" /> <img src="negative2.png" alt="Negative Examples" /> <img src="negative3.png" alt="Negative Examples" /></td>
</tr>
</tbody>
</table>
## Color and Shape

### Performance of trained model
- Ability to correctly classify held-out test set

### Goal: classifiers associated with attribute keywords have strong predictive power (only)

<table>
<thead>
<tr>
<th>Color classifier denoted by “term”</th>
<th>Ground truth</th>
<th>yellow</th>
<th>red</th>
<th>green</th>
<th>white</th>
<th>orange</th>
</tr>
</thead>
<tbody>
<tr>
<td>“yellow”</td>
<td></td>
<td>0.93</td>
<td>0.20</td>
<td>0.37</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>“building”</td>
<td></td>
<td>0.09</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>“red”</td>
<td></td>
<td>0.00</td>
<td>0.89</td>
<td>0.05</td>
<td>0.16</td>
<td>0.35</td>
</tr>
<tr>
<td>“green”</td>
<td></td>
<td>0.27</td>
<td>0.00</td>
<td>0.89</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>“tomato”</td>
<td></td>
<td>0.24</td>
<td>0.94</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>“white”</td>
<td></td>
<td>0.06</td>
<td>0.68</td>
<td>0.55</td>
<td>0.85</td>
<td>0.73</td>
</tr>
<tr>
<td>“orange”</td>
<td></td>
<td>0.50</td>
<td>0.93</td>
<td>0.21</td>
<td>0.26</td>
<td>0.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shape classifier denoted by “term”</th>
<th>Ground truth</th>
<th>cube</th>
<th>cylinder</th>
<th>sphere</th>
<th>arch</th>
<th>triangle</th>
</tr>
</thead>
<tbody>
<tr>
<td>“cylinder”</td>
<td></td>
<td>0.32</td>
<td>0.87</td>
<td>0.06</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>“rectangular”</td>
<td></td>
<td>0.82</td>
<td>0.43</td>
<td>0.51</td>
<td>0.78</td>
<td>0.30</td>
</tr>
<tr>
<td>“circle”</td>
<td></td>
<td>0.25</td>
<td>0.25</td>
<td>0.75</td>
<td>0.26</td>
<td>0.21</td>
</tr>
<tr>
<td>“archshaped”</td>
<td></td>
<td>0.29</td>
<td>0.27</td>
<td>0.12</td>
<td>0.82</td>
<td>0.33</td>
</tr>
<tr>
<td>“triangle”</td>
<td></td>
<td>0.54</td>
<td>0.60</td>
<td>0.52</td>
<td>0.31</td>
<td>0.82</td>
</tr>
</tbody>
</table>
Results: Object identification

- Object classification:
  - Possibility of learning more complex concepts
  - Good performance on interacting problem

<table>
<thead>
<tr>
<th>Object classifier denoted by “term”</th>
<th>corn</th>
<th>semi-cylinder</th>
<th>banana</th>
<th>eggplant</th>
<th>tomato</th>
</tr>
</thead>
<tbody>
<tr>
<td>“corn”</td>
<td>0.92</td>
<td>0.01</td>
<td>0.77</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>“building”</td>
<td>0.08</td>
<td>0.61</td>
<td>0.30</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>“banana”</td>
<td>0.00</td>
<td>0.15</td>
<td>1.00</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>“tomato”</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.94</td>
</tr>
<tr>
<td>“wedge”</td>
<td>0.49</td>
<td>0.30</td>
<td>0.00</td>
<td>0.43</td>
<td>0.00</td>
</tr>
<tr>
<td>“eggplant”</td>
<td>0.26</td>
<td>0.24</td>
<td>0.01</td>
<td>0.84</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Future Work

- A thorough evaluation in positive and negative term selection
  - Use Amazon Mechanical Turk

- Comparison of model with a traditional base model

- Evaluate the model in a more ‘real world’ problem
  - A more varied set of objects.
  - Additional kinds of classifiers.
  - Complex visual classification tasks.
Conclusion

- Semantic representations of their perceived environments
- Discovered ground truth labels
- Document similarity metrics in negative example selection
  - Efficient in unprompted human interaction scenario
  - Effective for grounded language acquisition tasks

Thank you! Questions?