Unit 6: Natural Language Processing

Dave Abel

March 11th, 2016
Reminder: Midterm 3/21

• One Week from Monday.

• You can bring a single sided notecard.

• Will cover material from Unit 0 to today (not Unit 7, which is next week)

• Similar to homework. Only material from lab/lecture/homework.

• Review Session Saturday 3/19, 4:30-6:00pm. Room TBD.

• **First 5 minutes of each class next week I’ll answer review questions: start bringing Questions for Monday.**
AI Announcement

AlphaGo 2-0 against Lee Se-Dol
Language
Language
Language

1. Thought
“Hey, do you want to go…”

1. Thought
2. Words
Language

1. Thought
2. Words

“Hey, do you want to go…”
1. Thought
2. Words
3. Coordinated nerve control

“Hey, do you want to go...”

Language

Tongue, Lips

Larynx

Lungs
1. Thought
2. Words
3. Coordinated nerve control
4. Sound

“Hey, do you want to go...”
1. Thought
2. Words
3. Coordinated nerve control
4. Sound

Language

“Hey, do you want to go…”

(Changes in air pressure)

Tongue, Lips

Larynx

Lungs
1. Thought
2. Words
3. Coordinated nerve control
4. Sound

“Hey, do you want to go...”

(Changes in air pressure)

Tongue, Lips
Larynx
Lungs

Language
Language

1. Thought
2. Words
3. Coordinated nerve control
4. Sound
5. Back to words

“Hey, do you want to go…”

(Changes in air pressure)

Words

Lungs

Tongue, Lips

Larynx
Language

1. Thought
2. Words
3. Coordinated nerve control
4. Sound
5. Back to words
6. Back to thought

“Hey, do you want to go…”

(Changes in air pressure)

Tongue, Lips
Larynx
Lungs

Words
Natural Language Processing

1. Thought
2. Words
3. Coordinated nerve control
4. Sound
5. Back to words
6. Back to thought
Natural Language Processing

IN:  1. Thought
OUT: 2. Words

3. Coordinated nerve control
4. Sound
5. Back to words
6. Back to thought
Natural Language Processing

1. Thought

INPUT:  2. Words

OUTPUT:  3. Coordinated nerve control

5. Back to words

4. Sound

6. Back to thought
Natural Language Processing

1. Thought
2. Words
3. Coordinated nerve control
4. Sound
5. Back to words
6. Back to thought
Natural Language Processing

1. Thought

2. Words

3. Coordinated nerve control

IN: 4. Sound

OUT: 5. Back to words

6. Back to thought
Natural Language Processing

1. Thought
2. Words
3. Coordinated nerve control
4. Sound

IN: 5. Back to words
OUT: 6. Back to thought
Natural Language Processing: Problems

- Speech Synthesis (*Words to Sound*)
- Speech Understanding (*Sound to Words*)
- Parsing, (*Words to Structured Sentences*)
- Semantic Parsing (*Words to Meaningful Sentences*)
- Language Modeling
- Sentiment Analysis
- Dialogue Systems
Natural Language Processing: Problems

› Speech Synthesis \((\text{Words to Sound})\)

› Speech Understanding \((\text{Sound to Words})\)

› Parsing, \((\text{Words to Structured Sentences})\)

› Semantic Parsing \((\text{Words to Meaningful Sentences})\)

› Language Modeling

› Sentiment Analysis (and more…)

› Dialogue Systems
Speech Synthesis: Humans

- Air: comes from your lungs!
- Forcing the air into particular shape: larynx, tongue, and lips do this!
- Try it: “Pit” vs. “Bit”
- Try it: “Cat” vs “Bat”
- Try it: “Lava”
Speech Synthesis: Computer

- Problem Specification:
  - INPUT: Words (text)
  - OUTPUT: Sound that corresponds to those words.
- One idea: piece together human speech…
Speech Synthesis

- One idea: piece together human speech…

https://www.youtube.com/watch?v=CgX4uJSj00Y
Speech Synthesis: Computer

‣ Problem Specification:

‣ INPUT: Words (text)

‣ OUTPUT: Sound that corresponds to those words.

‣ One idea: piece together human speech…

‣ Other idea: Machine Learning!
Speech Recognition: Human

› Our ears/brain somehow turn this:

› Into this:
Speech Recognition: Human

- Our ears/brain somehow turn this:

- Into this:

“Hello, my name is David”
Speech Recognition: Human

- Our ears/brain somehow turn this:

- Into this:

“Hello, my name is David”

0,0,0,0,44,87,99,88,84,29,44,22,11,8
Speech Recognition: Human

Q: What sound is this?
Speech Recognition: Human

Q: What sound is this?

“Kitten”
Speech Recognition: Human

Gaps in words aren’t the only “silence”

“Kitten”

Q: What sound is this?
Words to Thought

- Once we have some words, in some sense, we still have to put meaning to it..
Words to Thought

- Once we have some words, in some sense, we still have to put meaning to it..

- How about… (phonetic ambiguity)
Garden Path Sentences
Garden Path Sentences

“The old man the boat”
Garden Path Sentences

“The old man the boat”

Those who are old, man the boat.
Garden Path Sentences

“The old man the boat”

Those who are old, man the boat.

“Time Flies Like an Arrow”
Garden Path Sentences

“The old man the boat”

Those who are old, man the boat.

“Time Flies Like an Arrow”

“Fruit Flies Like a Banana”

Q: Is Time an adjective or a noun?
Garden Path Sentences

“The old man the boat”

Those who are old, man the boat.

“Time Flies Like an Arrow”

“Fruit Flies Like a Banana”

Q: Is Time an adjective or a noun?
A: Solving this problem is calling “parsing”
Dialogue Systems

- Chinese Room…

[Searle 1980]
Dialogue Systems

- Chinese Room…
- ELIZA [Weizenbaum ’96]

[Searle 1980]
Dialogue Systems

- Chinese Room…
- **ELIZA** [Weizenbaum ’96]
- **Cleverbot**

[Weizenbaum 1967]
[Searle 1980]
Dialogue Systems

‣ Chinese Room…
‣ **ELIZA** [Weizenbaum ’96]
‣ **Cleverbot**
‣ **SHRDLU**
SHRDLU

- Person: PICK UP A BIG RED BLOCK.
- Computer: OK. (does it)
- Person: GRASP THE PYRAMID.
- Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
- Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND
- PUT IT INTO THE BOX.
- Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
- Computer: OK. (does it)

[Winograd ’70]
Language Models

- We build a *model* that represents how we write language.

- We can build a specific model for certain types of language:
  - Text messaging auto complete
  - Scientific Paper Writing
  - Post Modernism Generator

- We can model how certain *authors* write!
Language Models

- Here’s the idea:
  - Given some text, written by a given author, count the number of times each word appears (usually ignoring silly words, like “the”).
  - Suppose we do this for a Roald Dahl novel (Fantastic Mr. Fox) and for a Lewis Carrol novel (Alice in Wonderland).
  - Given the word, “fantastic”, we ask, in which text did this appear more often?
Language Models: Counting

**Fantastic Mr. Fox**

<table>
<thead>
<tr>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
</tr>
</tbody>
</table>

**Alice in Wonderland**

<table>
<thead>
<tr>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>28</td>
</tr>
<tr>
<td>tree</td>
<td>13</td>
</tr>
<tr>
<td>fox</td>
<td>1</td>
</tr>
<tr>
<td>hit</td>
<td>1</td>
</tr>
<tr>
<td>hand</td>
<td>11</td>
</tr>
</tbody>
</table>

word  count  

word  count
Language Models: Counting

<table>
<thead>
<tr>
<th>Fantastic Mr. Fox</th>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alice in Wonderland</th>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>tree</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>fox</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>hit</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>hand</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

Q: Which author is more likely to use the word “fox”? 

The word count for “fox” in Fantastic Mr. Fox is 87, while in Alice in Wonderland it is 1. Therefore, Fantastic Mr. Fox is more likely to use the word “fox.”
Q: Which author is more likely to use the word “fox”?

Language Models: Counting

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>28</td>
</tr>
<tr>
<td>tree</td>
<td>13</td>
</tr>
<tr>
<td>fox</td>
<td>1</td>
</tr>
<tr>
<td>hit</td>
<td>1</td>
</tr>
<tr>
<td>hand</td>
<td>11</td>
</tr>
</tbody>
</table>

Fantastic Mr. Fox

Alice in Wonderland
Q: Which author is more likely to use the word “fox”? 

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>fox</td>
<td>17</td>
</tr>
</tbody>
</table>
Counting Problem

Dave’s Haiku

- fox

word count

Idea: Also need to take into account how big the text is!

Q: Which author is more likely to use the word “fox”? 
Probability An Author Will Write a Word

To do that, we use the *probability* the author will write a word:

\[
\text{Probability(word)} = \frac{\# \text{ times the author used the word}}{\text{Total # words the author has used}}
\]
To do that, we use the probability the author will write a word:

\[
\text{Probability(word)} = \frac{\text{# times the author used the word}}{\text{Total # words the author has used}}
\]

Not unique words, \textit{total} words
Q: What is the probability each text used the word fox?

<table>
<thead>
<tr>
<th>Fantastic Mr. Fox</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dave’s Haiku</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>fox</td>
<td>17</td>
</tr>
</tbody>
</table>
Probability!

Probability(word) = 

\[
\frac{\text{# times the author used the word}}{\text{Total # words the author has used}}
\]

Dave’s Haiku

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>fox</td>
<td>17</td>
</tr>
</tbody>
</table>
Probability!

Probability(fox) = \frac{\text{# times the author used "fox"}}{\text{Total # words the author has used}}

Q: What is the probability each text used the word fox?
Q: What is the probability each text used the word fox?

Probability!(fox) = \[
\frac{\text{17}}{\text{17}}
\]
Q: What is the probability each text used the word fox?

Probability(fox) = \[
\frac{17}{17} = 1.0
\]
Q: What is the probability each text used the word fox?
Probability!

Q: What is the probability each text used the word fox?

**Fantastic Mr. Fox**

<table>
<thead>
<tr>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
</tr>
</tbody>
</table>

**Probability(word)** = \[
\frac{\text{# times the author used the word}}{\text{Total # words the author has used}}
\]
Q: What is the probability each text used the word fox?

Fantastic Mr. Fox

<table>
<thead>
<tr>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
</tr>
</tbody>
</table>

Probability(fox) = \frac{\text{# times the author used “fox”}}{\text{Total # words the author has used}}
Q: What is the probability each text used the word fox?

Fantastic Mr. Fox

<table>
<thead>
<tr>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\text{Probability(fox)} = \frac{87}{133}
\]
Q: What is the probability each text used the word fox?

Probability (fox) = \frac{87}{133} = .65
Q: What is the probability each text used the word fox?

Fantastic Mr. Fox

<table>
<thead>
<tr>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>2</td>
</tr>
<tr>
<td>tree</td>
<td>31</td>
</tr>
<tr>
<td>fox</td>
<td>87</td>
</tr>
<tr>
<td>hit</td>
<td>3</td>
</tr>
<tr>
<td>hand</td>
<td>1</td>
</tr>
</tbody>
</table>

Word count = 65

Dave’s Haiku

<table>
<thead>
<tr>
<th>word</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>fox</td>
<td>17</td>
</tr>
</tbody>
</table>

Word count = 17

Probability of fox = \( \frac{65}{17} = 3.83 \)
Q: What is the probability each text used the word fox?

**Fantastic Mr. Fox**
- cat: 2
- tree: 31
- fox: 87
- hit: 3
- hand: 1

Word count = 118

Probability = \( \frac{87}{118} = 0.65 \)

**Dave’s Haiku**
- fox: 17

Word count = 31

Probability = 1.0
Probability!

Q: How about a full sentence?
Probability!

Q: How about a full sentence?

A: We just multiply each word’s probabilities together
Reflection

› Natural Language Processing is about trying to understand every facet of language, using computation and algorithms.

› It’s *crazy* hard.

› But super cool!
Next Week

Q: What, if anything, is out here?

A: The Halting Problem

Things that can be computed, period.

Things a regular computer can compute before the sun goes supernova

Things a domino computer could compute before the sun goes supernova

Halting
Next Week

The biggest unanswered question in computer science!