<table>
<thead>
<tr>
<th>Lecture #</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9 Sept</td>
<td>Intro Lecture</td>
</tr>
<tr>
<td>1</td>
<td>13 Sept</td>
<td>Tasks and Metrics</td>
</tr>
<tr>
<td>2</td>
<td>15 Sept</td>
<td>Text Classifiers Part 1: Models</td>
</tr>
<tr>
<td>3</td>
<td>20 Sept</td>
<td>Text Classifiers Part 2: Features</td>
</tr>
<tr>
<td>4</td>
<td>22 Sept</td>
<td>Tokenization</td>
</tr>
<tr>
<td>5</td>
<td>27 Sept</td>
<td>Distributional Hypothesis</td>
</tr>
<tr>
<td></td>
<td>29 Sept</td>
<td>Guest Lecture (Charlie Lovering): Computational Pragmatics</td>
</tr>
<tr>
<td>6</td>
<td>4 Oct</td>
<td>Topic Modeling</td>
</tr>
<tr>
<td>7</td>
<td>6 Oct</td>
<td>Deep Learning 101</td>
</tr>
<tr>
<td>8</td>
<td>11 Oct</td>
<td>Language Modeling Part 1: Ngram Models</td>
</tr>
<tr>
<td>9</td>
<td>13 Oct</td>
<td>Language Modeling Part 2: Neural Architectures</td>
</tr>
<tr>
<td>10</td>
<td>18 Oct</td>
<td>Language Modeling Part 3: Pretraining and Finetuning</td>
</tr>
<tr>
<td>11</td>
<td>20 Oct</td>
<td>Guest Lecture (Julia Netter): Risks of LM Misuse</td>
</tr>
<tr>
<td>12</td>
<td>25 Oct</td>
<td>Statistical Machine Translation</td>
</tr>
<tr>
<td>13</td>
<td>27 Oct</td>
<td>Neural Machine Translation</td>
</tr>
<tr>
<td>14</td>
<td>1 Nov</td>
<td>Guest Lecture (Dan Garrette): Technical and Societal Challenges of Multilingual NLP</td>
</tr>
<tr>
<td>Lecture #</td>
<td>Date</td>
<td>Topic</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>0</td>
<td>9 Sept</td>
<td>Intro Lecture</td>
</tr>
<tr>
<td>1</td>
<td>13 Sept</td>
<td>Tasks and Metrics</td>
</tr>
<tr>
<td>2</td>
<td>15 Sept</td>
<td>Text Classifiers Part 1: Models</td>
</tr>
<tr>
<td>3</td>
<td>20 Sept</td>
<td>Text Classifiers Part 2: Features</td>
</tr>
<tr>
<td>4</td>
<td>22 Sept</td>
<td>Tokenization</td>
</tr>
<tr>
<td>5</td>
<td>27 Sept</td>
<td>Distribution Hypothesis</td>
</tr>
<tr>
<td>6</td>
<td>29 Sept</td>
<td>Guest Lecture (Charlie Lovering): Computational Pragmatics</td>
</tr>
<tr>
<td>7</td>
<td>4 Oct</td>
<td>Topic Modeling</td>
</tr>
<tr>
<td>8</td>
<td>6 Oct</td>
<td>Deep Learning 101</td>
</tr>
<tr>
<td>9</td>
<td>11 Oct</td>
<td>Language Modeling Part 1: Ngram Models</td>
</tr>
<tr>
<td>10</td>
<td>13 Oct</td>
<td>Language Modeling Part 2: Neural Architectures</td>
</tr>
<tr>
<td>11</td>
<td>18 Oct</td>
<td>Language Modeling Part 3: Pretraining and Finetuning</td>
</tr>
<tr>
<td>12</td>
<td>20 Oct</td>
<td>Guest Lecture (Julia Netter): Risks of LM Misuse</td>
</tr>
<tr>
<td>13</td>
<td>25 Oct</td>
<td>Statistical Machine Translation</td>
</tr>
<tr>
<td>14</td>
<td>27 Oct</td>
<td>Neural Machine Translation</td>
</tr>
<tr>
<td></td>
<td>1 Nov</td>
<td>Guest Lecture (Dan Garrette): Technical and Societal Challenges of Multilingual NLP</td>
</tr>
</tbody>
</table>
milkyway in a glass bottle, 4K, unreal engine, octane render

Left: https://www.midjourney.com/showcase/
Model: Midjourney
Author: Deeds
Standard prompting

Input:
Q: Shawn has five toys. For Christmas, he got two toys each from his mom and dad. How many toys does he have now?
A: The answer is 9.

... 

Q: John takes care of 10 dogs. Each dog takes .5 hours a day to walk and take care of their business. How many hours a week does he spend taking care of dogs?
A:

Model output: The answer is 50. ❌

Chain of thought prompting

Input:
Q: Shawn has five toys. For Christmas, he got two toys each from his mom and dad. How many toys does he have now?
A: Shawn started with 5 toys. If he got 2 toys each from his mom and dad, then that is 4 more toys. $5 + 4 = 9$. The answer is 9.

... 

Q: John takes care of 10 dogs. Each dog takes .5 hours a day to walk and take care of their business. How many hours a week does he spend taking care of dogs?
A:

Model output: John takes care of 10 dogs. Each dog takes .5 hours a day to walk and take care of their business. So that is $10 \times .5 = 5$ hours a day. 5 hours a day $\times 7$ days a week $= 35$ hours a week. The answer is 35 hours a week. ✔️
Standard prompting

Input:
Q: Shawn has five toys. For Christmas, he got two toys each from his mom and dad. How many toys does he have now?
A: The answer is 9.

Q: John takes care of 10 dogs. Each dog takes .5 hours a day to walk and take care of their business. How many hours a week does he spend taking care of dogs?
A:

Model output: The answer is 50. ❌

Chain of thought prompting

Input:
Q: Shawn has five toys. For Christmas, he got two toys each from his mom and dad. How many toys does he have now?
A: Shawn started with 5 toys. If he got 2 toys each from his mom and dad, then that is 4 more toys. 5 + 4 = 9. The answer is 9.

Q: John takes care of 10 dogs. Each dog takes .5 hours a day to walk and take care of their business. How many hours a week does he spend taking care of dogs?
A:

Model output: John takes care of 10 dogs. Each dog takes .5 hours a day to walk and take care of their business. So that is 10 x .5 = 5 hours a day. 5 hours a day x 7 days a week = 35 hours a week. The answer is 35 hours a week. ✅

Left: https://www.midjourney.com/showcase/
Model: Midjourney
Author: Deeds

Model: T5-like, gpt3-like models (google internal)
Pragmatics
Rational Speech Acts

Very-low Data
I went to the bank.
I went to the bank.

https://en.wikipedia.org/wiki/Pragmatics
I went to the bank.

https://en.wikipedia.org/wiki/Pragmatics
You have a green light.
You have a green light.
You have a green light.
Scales

The big mouse; the small elephant

Hyperbole

Damn, those shoes must cost a million dollars.

Metaphor

He was a beast out there on the court.

Sarcasm

Digging this hole is so much fun.

https://en.wikipedia.org/wiki/Pragmatics
Reference Games
What would you say to get a “listener” pick the selected item?
What would you pick?

Orange.
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]
\[ P_{\text{Literal}}(r | m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]
If message applies to a given referent
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

Prior of the referent.
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

Normalizing across referents in context
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]
\[ P_{\text{Literal}}(\square \mid \text{“orange”}) = \frac{\text{Eval}(\square, \text{“orange”}) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]
\[ P_{\text{Literal}}(\square \mid \text{“orange”}) = \frac{\text{Eval}(\square, \text{“orange”}) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ = \frac{0 \times \frac{1}{3}}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ = 0. \]

<table>
<thead>
<tr>
<th>Eval</th>
<th>0</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>“orange”</td>
<td>(0)</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\begin{array}{c|c|c|c|c|c}
\text{Eval} & \square & \square & \square & \text{P} & \text{P}_{\text{Literal}}(\ast \mid \text{“orange”}) = \\
\hline
\text{“orange”} & 0 & 1 & 1 & \frac{1}{3} & 0 \\
\hline
\end{array}
\]
\[ P_{\text{Literal}}(\square | \text{“orange”}) = \frac{\text{Eval}(\square, \text{“orange”}) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]
\[ = \frac{1 \times \frac{1}{3}}{1 \times \frac{1}{3} + 1 \times \frac{1}{3}} \]
\[ = 0.5 \]
\[
P_{\text{Literal}}(\square \mid \text{“orange”}) = P_{\text{Literal}}(\bullet \mid \text{“orange”})
\]
\[
= \frac{\text{Eval}(\bullet, \text{“orange”}) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}
\]
\[
= \frac{1 \times \frac{1}{3}}{1 \times \frac{1}{3} + 1 \times \frac{1}{3}}
\]
\[
= 0.5
\]
\[
P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}
\]

\[
P_{\text{Literal}}(\ast \mid \text{“circle”})
\]
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ P_{\text{Literal}}(\ast \mid \text{“circle”}) \]
What’s next?
Gricean Principles

Quality: Say what’s true
Quantity: Be informative; don’t say more than necessary
Relevance: Be relevant
Manner: Don’t be ambiguous, be brief, ...
“What would I say if I were them and wanted me to pick XYZ?”
What would I say if I were them and wanted me to pick XYZ?“

While not being ambiguous, and saying only what is needed.
\begin{align*}
P_{\text{Literal}}(r \mid m) &= \text{P of a referent given a message, understood literally} \\
\end{align*}

\begin{align*}
P_S(m \mid r) &= \text{P of a message, assuming the message will be heard by a literal listener.} \\
\end{align*}

\begin{align*}
P_L(r \mid m) &= \text{P of a referent, assuming the message was sent by a pragmatic speaker.} \\
\end{align*}
\[
P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}
\]

\[
P_{S}(m \mid r) = \text{P of a message, assuming the message will be heard by a literal listener.}
\]

\[
P_{L}(r \mid m) = \text{P of a referent, assuming the message was sent by a pragmatic speaker.}
\]
$$P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}$$

$$P_S(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')}$$

$$P_L(r \mid m) = \text{P of a referent, assuming the message was sent by a pragmatic speaker.}$$
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ P_{S}(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \]

\[ P_{L}(r \mid m) = \frac{P_{S}(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_{S}(m \mid r) \times P(r')} \]
What would you pick?

Orange.
\[
\begin{align*}
P_{\text{Literal}}(r \mid m) &= \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \\
&= \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \\
&= \frac{P_{S}(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_{S}(m \mid r') \times P(r')}
\end{align*}
\]
\[ P_{\text{Literal}}(\square \mid \text{“orange”}) = P_{\text{Literal}}(\bullet \mid \text{“orange”}) \]

\[ = \frac{\text{Eval}(\bullet, \text{“orange”}) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ = \frac{1 \times \frac{1}{3}}{1 \times \frac{1}{3} + 1 \times \frac{1}{3}} \]

\[ = 0.5 \]
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ P_S(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \]

\[ P_L(r \mid m) = \frac{P_S(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_S(m \mid r') \times P(r')} \]
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ P_{S}(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \]

\[ P_{L}(r \mid m) = \frac{P_{S}(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_{S}(m \mid r') \times P(r')} \]
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ P_S(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \]

\[ P_L(r \mid m) = \frac{P_S(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_S(m \mid r') \times P(r')} \]

\[ P_S(\text{“orange”} \mid r) = \frac{P_{\text{Literal}}(r \mid \text{“orange”})}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \]

\[ P_S(\text{“orange”} \mid \square) = \frac{P_{\text{Literal}}(\square \mid \text{“orange”})}{P_{\text{Lit}}(\square \mid \text{“blue”}) + P_{\text{Lit}}(\square \mid \text{“orange”}) + P_{\text{Lit}}(\square \mid \text{“square”}) + P_{\text{Lit}}(\square \mid \text{“circle”})} \]

\[ = \frac{0}{1 + 0 + 0.5 + 0} = 0. \]

\[
\begin{array}{|c|c|c|c|}
\hline
& \text{“blue”} & \text{“orange”} & \text{“square”} & \text{“circle”} \\
\hline
\text{“blue”} & 1 & 0 & 0.5 & 0 \\
\hline
\text{“orange”} & 0 & 0.5 & 0.5 & 0 \\
\hline
\text{“square”} & 0 & 0.5 & 0 & 1 \\
\hline
\end{array}
\]
\[
P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}
\]

\[
P_{S}(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \quad \frac{P_{\text{Literal}}(r \mid m)}{P_{\text{Literal}}(r \mid m')}
\]

\[
P_{L}(r \mid m) = \frac{P_{S}(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_{S}(m \mid r') \times P(r')}
\]

<table>
<thead>
<tr>
<th>P_{\text{Literal}}(r \mid m)</th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P_{S}(m \mid r)</th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“blue”</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>“orange”</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>“square”</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>“circle”</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ P_S(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \]

\[ P_L(r \mid m) = \frac{P_S(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_S(m \mid r') \times P(r')} \]

\[ P_S(\text{“orange”} \mid r) = \frac{P_{\text{Literal}}(r \mid \text{“orange”})}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \]

\[ P_S(\text{“orange”} \mid □) = \frac{P_{\text{Literal}}(□ \mid \text{“orange”})}{P_{\text{Lit}}(□ \mid \text{“blue”}) + P_{\text{Lit}}(□ \mid \text{“orange”}) + P_{\text{Lit}}(□ \mid \text{“square”}) + P_{\text{Lit}}(□ \mid \text{“circle”})} \]

\[ = \frac{0.5}{0 + 0.5 + 0.5 + 0} = 0.5 \]
\[
P_{\text{Literal}}(r | m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}
\]

\[
P_S(m | r) = \frac{P_{\text{Literal}}(r | m) \times P_S(m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r | m') \times P_S(m')}
\]

\[
P_L(r | m) = \frac{P_S(m | r) \times P(r)}{\sum_{r' \in \text{Context}} P_S(m | r') \times P(r')}
\]
\[
P_{\text{Literal}}(r | m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}
\]

\[
P_{S}(m | r) = \frac{P_{\text{Literal}}(r | m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r | m')}
\]

\[
P_{L}(r | m) = \frac{P_{S}(m | r) \times P(r)}{\sum_{r' \in \text{Context}} P_{S}(m | r') \times P(r')}
\]

\[
P_{S}(\text{“orange”} | r) = \frac{P_{\text{Literal}}(r | \text{“orange”})}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r | m')}
\]

\[
P_{S}(\text{“orange”} | \bigcirc) = \frac{P_{\text{Literal}}(\bigcirc | \text{“orange”})}{P_{\text{Lit}}(\bigcirc | \text{“blue”}) + P_{\text{Lit}}(\bigcirc | \text{“orange”}) + P_{\text{Lit}}(\bigcirc | \text{“square”}) + P_{\text{Lit}}(\bigcirc | \text{“circle”})}
\]

\[
= \frac{0.5}{0 + 0.5 + 0 + 1} = 0.33
\]
\[ P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \]

\[ P_S(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \]

- **“blue”**
  - \[ P_{\text{Literal}}(r \mid m) = \]
  - \[
  \begin{array}{cccc}
  P_{\text{Literal}}(r \mid m) & \text{“blue”} & \text{“orange”} & \text{“square”} & \text{“circle”} \\
  \hline
  \text{blue} & 1 & 0 & 0.5 & 0 \\
  \text{orange} & 0 & 0.5 & 0.5 & 0 \\
  \text{square} & 0 & 0.5 & 0 & 1 \\
  \text{circle} & ? & ? & ? & ?
  \end{array}
  \]

- **“orange”**
  - \[ P_{\text{Literal}}(r \mid m) = \]
  - \[
  \begin{array}{cccc}
  P_{\text{Literal}}(r \mid m) & \text{“blue”} & \text{“orange”} & \text{“square”} & \text{“circle”} \\
  \hline
  \text{blue} & 1 & 0 & 0.5 & 0 \\
  \text{orange} & 0 & 0.5 & 0.5 & 0 \\
  \text{square} & 0 & 0.5 & 0 & 1 \\
  \text{circle} & ? & ? & ? & ?
  \end{array}
  \]

- **“square”**
  - \[ P_{\text{Literal}}(r \mid m) = \]
  - \[
  \begin{array}{cccc}
  P_{\text{Literal}}(r \mid m) & \text{“blue”} & \text{“orange”} & \text{“square”} & \text{“circle”} \\
  \hline
  \text{blue} & 1 & 0 & 0.5 & 0 \\
  \text{orange} & 0 & 0.5 & 0.5 & 0 \\
  \text{square} & 0 & 0.5 & 0 & 1 \\
  \text{circle} & ? & ? & ? & ?
  \end{array}
  \]

- **“circle”**
  - \[ P_{\text{Literal}}(r \mid m) = \]
  - \[
  \begin{array}{cccc}
  P_{\text{Literal}}(r \mid m) & \text{“blue”} & \text{“orange”} & \text{“square”} & \text{“circle”} \\
  \hline
  \text{blue} & 1 & 0 & 0.5 & 0 \\
  \text{orange} & 0 & 0.5 & 0.5 & 0 \\
  \text{square} & 0 & 0.5 & 0 & 1 \\
  \text{circle} & ? & ? & ? & ?
  \end{array}
  \]
\[
P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}
\]

\[
P_{S}(m \mid r) = \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')}
\]

A table showing the probabilities for different contexts:

<table>
<thead>
<tr>
<th>Context</th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“blue”</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>“orange”</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>“square”</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>“circle”</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
</tbody>
</table>

A separate table with missing values indicated by ?.
\[
\begin{align*}
P_{\text{Literal}}(r \mid m) &= \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')} \\
\Pr(m \mid r) &= \frac{P_{\text{Literal}}(r \mid m)}{\sum_{m' \in \text{Messages}} P_{\text{Literal}}(r \mid m')} \\
P_{L}(r \mid m) &= \frac{P_{\text{S}}(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_{\text{S}}(m \mid r') \times P(r')}
\end{align*}
\]

<table>
<thead>
<tr>
<th>P_{\text{Literal}}(r \mid m)</th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P_{\text{S}}(m \mid r)</th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“blue”</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>“orange”</td>
<td>0</td>
<td>0.5</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>“square”</td>
<td>0.33</td>
<td>0.5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>“circle”</td>
<td>0</td>
<td>0</td>
<td>0.67</td>
<td></td>
</tr>
</tbody>
</table>
\[
P_L(r \mid m) = \frac{P_S(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_S(m \mid r) \times P(r')}
\]

\[
P_L(\text{“orange”}) = \frac{P_S(\text{“orange”} \mid \text{orange}) \times P(\text{orange})}{\sum_{r' \in \text{Context}} P_S(\text{“orange”} \mid r') \times P(r')}
\]

\[
= \frac{0.5 \times 0.33}{0 \times 0.33 + 0.5 \times 0.33 + 0.33 \times 0.33}
= \frac{1/2}{1/2 + 1/3}
= \frac{3}{5}.
\]
\[
P_L(r \mid m) = \frac{P_S(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_S(m \mid r') \times P(r')}
\]

\[
P_L(\text{“orange”}) = \frac{P_S(\text{“orange”}) \mid \square \times P(\square)}{\sum_{r' \in \text{Context}} P_S(\text{“orange”}) \mid r' \times P(r')}
\]

\[
= \frac{0.5 \times 0.33}{0 \times 0.33 + 0.5 \times 0.33 + 0.33 \times 0.33} = \frac{1/2}{1/2 + 1/3} = \frac{3}{5}
\]

\[
P_L(\text{“orange”}) = \frac{P_S(\text{“orange”}) \mid \bullet \times P(\bullet)}{\sum_{r' \in \text{Context}} P_S(\text{“orange”}) \mid r' \times P(r')}
\]

\[
= \frac{0.33 \times 0.33}{0 \times 0.33 + 0.5 \times 0.33 + 0.33 \times 0.33} = \frac{1/3}{1/2 + 1/3} = \frac{2}{5}
\]
\[ P_L(\square | \text{“orange”}) = 0, \]
\[ P_L(\bigheartsuit | \text{“orange”}) = \frac{3}{5}, \]
\[ P_L(\bigdiamond | \text{“orange”}) = \frac{2}{5}. \]
<table>
<thead>
<tr>
<th>$P_{\text{Literal}}(r \mid m)$</th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Blue" /></td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Orange" /></td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Circle" /></td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$P_2(m \mid r)$</th>
<th><img src="image" alt="Blue" /></th>
<th><img src="image" alt="Orange" /></th>
<th><img src="image" alt="Square" /></th>
<th><img src="image" alt="Circle" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>“blue”</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>“orange”</td>
<td>0</td>
<td>0.5</td>
<td>0.33</td>
<td>0</td>
</tr>
<tr>
<td>“square”</td>
<td>0.33</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>“circle”</td>
<td>0</td>
<td>0</td>
<td>0.67</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$P_L(r \mid m)$</th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Blue" /></td>
<td>1</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Orange" /></td>
<td>0</td>
<td>0.6</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Circle" /></td>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
The diagram illustrates the process of transposing and normalizing a probability matrix. The matrices are as follows:

### $P_{\text{Literal}}(r \mid m)$

<table>
<thead>
<tr>
<th></th>
<th>&quot;blue&quot;</th>
<th>&quot;orange&quot;</th>
<th>&quot;square&quot;</th>
<th>&quot;circle&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;blue&quot;</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>&quot;orange&quot;</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>&quot;square&quot;</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### $P_{\text{Context}}(m \mid r)$

<table>
<thead>
<tr>
<th></th>
<th>&quot;blue&quot;</th>
<th>&quot;orange&quot;</th>
<th>&quot;square&quot;</th>
<th>&quot;circle&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;blue&quot;</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&quot;orange&quot;</td>
<td>0</td>
<td>0.5</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>&quot;square&quot;</td>
<td>0.33</td>
<td>0.5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>&quot;circle&quot;</td>
<td>0</td>
<td>0</td>
<td>0.67</td>
<td></td>
</tr>
</tbody>
</table>

### $P_{L}(r \mid m)$

<table>
<thead>
<tr>
<th></th>
<th>&quot;blue&quot;</th>
<th>&quot;orange&quot;</th>
<th>&quot;square&quot;</th>
<th>&quot;circle&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;blue&quot;</td>
<td>1</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>&quot;orange&quot;</td>
<td>0</td>
<td>0.6</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>&quot;square&quot;</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The process involves transposing and normalizing these matrices to update the probabilities.
### Table 1: $P_{\text{Literal}}(r|m)$

<table>
<thead>
<tr>
<th></th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“blue”</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>“orange”</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>“square”</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 2: $P_{\mathbb{S}}(m|r)$

<table>
<thead>
<tr>
<th></th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“blue”</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>“orange”</td>
<td>0</td>
<td>0.5</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>“square”</td>
<td>0.33</td>
<td>0.5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>“circle”</td>
<td>0</td>
<td>0</td>
<td>0.67</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: $P_{L}(r|m)$

<table>
<thead>
<tr>
<th></th>
<th>“blue”</th>
<th>“orange”</th>
<th>“square”</th>
<th>“circle”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“blue”</td>
<td>1</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>“orange”</td>
<td>0</td>
<td>0.6</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>“square”</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
What would you pick?

“Mushrooms”
\( P_{\text{Literal}}(r \mid m) \):

\[
\begin{array}{cccc}
\text{Pepperoni} & \text{mushroom} & \text{Cheese} & \text{Plain} \\
0.0 & 0.5 & 0.33 & 0.0 \\
0.0 & 0.5 & 0.33 & 0.0 \\
0.0 & 0.0 & 0.33 & 1.0 \\
\end{array}
\]

\( P_{S}(m \mid r) \):

<table>
<thead>
<tr>
<th></th>
<th>Pepperoni</th>
<th>mushroom</th>
<th>Cheese</th>
<th>Plain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peperoni</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mushroom</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheese</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plain</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\( P_{\text{Literal}}(r \mid m) \)

\( P_S(m \mid r) : \)
\[ P_{\text{Literal}}(r \mid m) \]

\[ P_{S}(m \mid r) : \]

- Pepperoni: 0
- mushroom: 0.6
- Cheese: 0.4
- Plain: 0

\[
\begin{align*}
0 & \rightarrow 0 + 0.5 + 0.33 + 0 = 0 \\
\frac{1}{2} & \rightarrow \frac{0 + 1/2 + 1/3 + 0}{1/2 + 1/3} = \frac{3}{5} = 0.6 \\
\frac{1}{3} & \rightarrow \frac{0 + 1/2 + 1/3 + 0}{1/2 + 1/3} = \frac{2}{5} = 0.4 \\
0 & \rightarrow 0 + 0.5 + 0.33 + 0 = 0
\end{align*}
\]
$P_{\text{Literal}}(r \mid m) = \frac{\text{Eval}(r, m) \times P(r)}{\sum_{r' \in \text{Context}} \text{Eval}(r', m) \times P(r')}$

$P_{S}(m \mid r) = \frac{\exp(\alpha(\log P_{\text{Literal}}(r \mid m) + C(m))))}{\sum_{m' \in \text{Messages}} \exp(\alpha(\log P_{\text{Literal}}(r \mid m') + C(m'))))}$

$P_{L}(r \mid m) = \frac{P_{S}(m \mid r) \times P(r)}{\sum_{r' \in \text{Context}} P_{S}(m \mid r) \times P(r')}$
Learning in the Rational Speech Acts Model

Will Monroe and Christopher Potts
Stanford University, California, U.S.A.
wmmonroe@cs.stanford.edu, cgpotts@stanford.edu

Abstract

The Rational Speech Acts (RSA) model treats language use as a recursive process in which probabilistic speaker and listener agents reason about each other's intentions to enrich the literal semantics of their language along broadly Gricean lines. RSA has been shown to capture many kinds of conversational implicature, but it has been criticized as an unrealistic model of speakers, and it has so far required the manual specification of a semantic lexicon, preventing its use in natural language processing applications that learn lexical knowledge from data. We address these concerns by showing how to define and optimize a trained statistical classifier that uses the intermediate agents of RSA as hidden layers of representation forming a non-linear activation function. This treatment opens up new application domains and new possibilities for learning effectively from data. We validate the model on a referential expression generation task, showing that the best performance is achieved by incorporating features approximating well-established insights about natural language generation into RSA.

Literal speaker. Learned RSA is built on top of a log-linear model, standard in the machine learning literature and widely applied to classification tasks [19, 27].

\[ S_0(m \mid t, c; \theta) \propto \exp(\theta^T \phi(t, m, c)) \]
There are probabilistic programming languages that support ‘neat’ implementations.

```python
@Marginal
def speaker(state):
    alpha = 1.
    with poutine.scale(scale=torch.tensor(alpha)):
        utterance = utterance_prior()
        pyro.sample("listener", literal_listener(utterance), obs=state)
    return utterance
```

Finally, we can define the pragmatic_listener, who infers which state is likely, given that the speaker chose a given utterance. Mathematically:

\[
P_L(s|u) \propto P_S(u|s)P(s)
\]

In code:

```python
@Marginal
def pragmatic_listener(utterance):
    state = state_prior()
    pyro.sample("speaker", speaker(state), obs=utterance)
    return state
```

https://pyro.ai/examples/index.html
Aside #4

Add more layers for Better performance.

Understanding the Rational Speech Act model

Variation across ppl
Pragmatics

Very-low Data
Pragmatics

Very-low Data
Bayesian Language of Thought
<table>
<thead>
<tr>
<th>1,2,3</th>
<th>1,2,3</th>
</tr>
</thead>
<tbody>
<tr>
<td>9,7,8</td>
<td>7,8,9</td>
</tr>
<tr>
<td>4,5</td>
<td>4</td>
</tr>
<tr>
<td>----------</td>
<td>----</td>
</tr>
<tr>
<td>4,5,9</td>
<td>4,9</td>
</tr>
<tr>
<td>4,5,9,2</td>
<td>4,9</td>
</tr>
<tr>
<td>4,5</td>
<td>4</td>
</tr>
<tr>
<td>-------------</td>
<td>----</td>
</tr>
<tr>
<td>4,5,9</td>
<td>4,9</td>
</tr>
<tr>
<td>4,5,9,2</td>
<td>4,9</td>
</tr>
<tr>
<td>4,5,9,3,2,16,9</td>
<td>4,9,16,9</td>
</tr>
</tbody>
</table>
Bayesian Language of Thought =

Set of primitives
Prior function
Inference Scheme
A Set of primitives

And
Or
Not
Pair
If
Recurse
Head
A Set of primitives

And
Or
Not
Pair
If
Recurse
Head
Tail
ForEach
Filter
Eq…
A Set of primitives

And 0..9
Or +
Not -
Pair /
If $\sqrt{\text{}$
Recurse ...
Head
Tail
ForEach
Filter
Eq…
A Set of primitives

And 0..9 Forward
Or + Backward
Not - Left
Pair / Right
If sqrt Goto
Recurse ... penup
Head pendown
Tail begin_fill
ForEach end_fill
Filter
Eq…
Prior: How likely is the $H$?

Inversely related to description length

How likely does $H$ make the observed $D$?
\[
H_1 = \text{ForEach}(x, x \times 2 \times 2 + 2 - 2 - x \times 2)
\]
\[
H_2 = \text{ForEach}(x, x \times 2)
\]
\[
P(H_1) < P(H_2)
\]
\[ H_1 = \text{ForEach}((x, i), \text{If}(\text{Eq}(i \mod 2, 0), x, \emptyset)) \]

\[ H_2 = \text{ForEach}(x, x \ast 2) \]

\[ P(D|H_1) > P(D|H_2) \]
Hypothesis Space

\[ P(H) P(D \mid H) \]
Hypothesis Space

P(H) P(D | H)

How to get there?
1. Make a random change to $H \rightarrow H'$
2. If $\frac{P(H')P(D|H')}{P(H)P(D|H)} > 1$, accept $H'$
3. Else, with some $p$, accept $H'$, or repeat.
Accept $H'$, set $H = H'$, repeat.
Accept $H'$ with $p$ setting $H = H'$, repeat.
Accept $H'$ with $p$ setting $H = H'$, repeat.

Note: In some schemes, $p$ is fixed, in some, it is simply the ratio.
Hastings Algorithm

1. Make a random change to $H \rightarrow H'$
2. If $\frac{P(H')P(D|H')}{P(H)P(D|H)} > 1$, accept $H'$
3. Else, with some $p$, accept $H'$, or repeat.
git clone https://github.com/piantado/Fleet

cd Fleet/Models/Sorting
module load eigen/3.4.0 mpi/openmpi_4.1.1_gcc_10.2_slurm20 gcc/10.2 clang/7.1.0
make
./main

<table>
<thead>
<tr>
<th>1,2,3</th>
<th></th>
<th>1,2,3</th>
</tr>
</thead>
<tbody>
<tr>
<td>9,7,8</td>
<td></td>
<td>7,8,9</td>
</tr>
</tbody>
</table>
1 minute

\[ \lambda x. \text{if}_s( \ (\text{reverse}(x) < x), \ \text{pair}( \ \text{tail}(x), \ \text{head}(x)), \ x) \]

<table>
<thead>
<tr>
<th>1,2,3</th>
<th>1,2,3</th>
</tr>
</thead>
<tbody>
<tr>
<td>9,7,8</td>
<td>7,8,9</td>
</tr>
</tbody>
</table>

3 minutes

\[ \lambda x. \text{tail}( \ \text{if}_s( \ (x < \text{tail}(x)), \ \text{tail}( \ \text{pair}( \ \text{reverse}(x), \ \text{head}( \ \text{recurse}(\text{reverse}(x))) ), \ \text{pair}(x, \text{head}(x)))) ) \]
<table>
<thead>
<tr>
<th>References / Things to read</th>
<th>References / Things to read</th>
</tr>
</thead>
<tbody>
<tr>
<td>For Pragmatics &amp; RSA</td>
<td>For Bayesian Language of Thought</td>
</tr>
</tbody>
</table>

- [https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0154854&type=printable](https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0154854&type=printable)

- One model for the learning of language
  [DreamCoder](https://github.com/piantado/Fleet)
  [Metropolis%E2%80%93Hastings_algorithm](https://en.wikipedia.org/wiki/Metropolis%E2%80%93Hastings_algorithm)
Thanks!