Language Models

Computational Linguistics: Jordan Boyd-Graber
University of Maryland
EXERCISE
Exercise

- Start with restaurant we had before
- Assume you see <s> b b a c </s>; add those counts to tables
- Compute probability of b following a (θ = 1.0, δ = 0.5)
- Compute the probability of a following b
- Compute probability of </s> following <s>
A busy night at the restaurant

<table>
<thead>
<tr>
<th>Unigram Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a³ b¹ c¹ &lt;/s&gt;¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&lt;s&gt; Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>a Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a² b¹ c¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;/s&gt;¹</td>
</tr>
</tbody>
</table>
A busy night at the restaurant

Unigram Restaurant

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>&lt;/s&gt;</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<s> Restaurant

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b Restaurant

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

c Restaurant

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;/s&gt;</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
A busy night at the restaurant

<table>
<thead>
<tr>
<th>Unigram Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a³ b² c¹ &lt;/s&gt;¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&lt;s&gt; Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a¹ b¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>a Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a² b¹ c¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;/s&gt;¹</td>
</tr>
</tbody>
</table>
A busy night at the restaurant

Unigram Restaurant

<s> Restaurant

b Restaurant

a Restaurant

c Restaurant
A busy night at the restaurant

Unigram Restaurant

\[
\begin{align*}
\text{a}^3 & \quad \text{b}^3 \quad \text{c}^1 \quad \langle /s \rangle^1
\end{align*}
\]

<s> Restaurant

\[
\begin{align*}
\text{a}^1 & \quad \text{b}^1
\end{align*}
\]

b Restaurant

\[
\begin{align*}
\text{a}^1 & \quad \text{b}^1
\end{align*}
\]

a Restaurant

\[
\begin{align*}
\text{a}^2 & \quad \text{b}^1 \quad \text{c}^1
\end{align*}
\]

c Restaurant

\[
\begin{align*}
\langle /s \rangle^1
\end{align*}
\]
A busy night at the restaurant

Unigram Restaurant

<s> Restaurant

b Restaurant

a Restaurant

<>/s> Restaurant
A busy night at the restaurant

<table>
<thead>
<tr>
<th>Unigram Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a³ b³ c¹ &lt;/s&gt;¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&lt;s&gt; Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a¹ b¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a² b¹</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>c²</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>a Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a² b¹ c²</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&lt;/s&gt;¹</th>
</tr>
</thead>
</table>
A busy night at the restaurant

Unigram Restaurant

```
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>&lt;/s&gt;</td>
</tr>
</tbody>
</table>
```

<s> Restaurant

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>
```

b Restaurant

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>
```

c Restaurant

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;/s&gt;</td>
<td></td>
</tr>
</tbody>
</table>
```
A busy night at the restaurant

<table>
<thead>
<tr>
<th>Unigram Restaurant</th>
<th>Restaurant</th>
<th>Restaurant</th>
<th>Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>a^{3} b^{3} c^{1}</td>
<td>a^{2} b^{1} c^{2}</td>
<td>a^{2} b^{1}</td>
<td>a^{2} b^{1} c^{2}</td>
</tr>
</tbody>
</table>

As you see more data, bottom restaurants do more work.
\[ \begin{align*}
\text{following } a \\
 \frac{1 - \delta}{\theta + 5} + \frac{\theta + 3\delta}{\theta + 5} p(b) & \quad (1) \\
\frac{1 - \delta}{\theta + 5} + \frac{\theta + 3\delta}{\theta + 5} \left( \frac{3 - \delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right) & \quad (2) \\
\end{align*} \]
\textbf{b following a}

\begin{align*}
&= \frac{1 - \delta}{\theta + 5} + \frac{\theta + 3\delta}{\theta + 5} p(b) \\
&= \frac{1 - \delta}{\theta + 5} + \frac{\theta + 3\delta}{\theta + 5} \left( \frac{3 - \delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right)
\end{align*}

(1) 

(2) 

(3)
\[ \begin{align*}
\frac{1 - \delta}{\theta + 5} + \frac{\theta + 3\delta}{\theta + 5} p(b) \\
\frac{1 - \delta}{\theta + 5} + \frac{\theta + 3\delta}{\theta + 5} \left( \frac{3 - \delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right) \\
0.23
\end{align*} \]
\[ a \text{ following } b \]

\[
\begin{align*}
    &= \frac{2 - \delta}{\theta + 3} + \frac{\theta + 2\delta}{\theta + 3} p(a) \\
    &= \frac{2 - \delta}{\theta + 3} + \frac{\theta + 2\delta}{\theta + 3} \left( \frac{3 - \delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right)
\end{align*}
\]
\[ a \text{ following } b \]

\[
\begin{align*}
2 - \delta &+ \frac{\theta + 2\delta}{\theta + 3} p(a) \\
&= \frac{2 - \delta}{\theta + 3} + \frac{\theta + 2\delta}{\theta + 3} \left( \frac{3 - \delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right)
\end{align*}
\]
\[ a \text{ following } b \]

\[
\begin{align*}
\text{(4)} & \quad = \frac{2-\delta}{\theta + 3} + \frac{\theta + 2\delta}{\theta + 3} p(a) \\
\text{(5)} & \quad = \frac{2-\delta}{\theta + 3} + \frac{\theta + 2\delta}{\theta + 3} \left( \frac{3-\delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right) \\
\text{(6)} & \quad 0.55
\end{align*}
\]
\[
\begin{align*}
\theta + 2\delta \frac{p(</s>)}{\theta + 2} \\
= \frac{\theta + 2\delta}{\theta + 2} \left( \frac{1 - \delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right)
\end{align*}
\]
\[ \frac{\theta + 2\delta}{\theta + 2} p(</s>) = \frac{\theta + 2\delta}{\theta + 2} \left( \frac{1 - \delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right) \]
\[ p(\langle s \rangle) = \frac{\theta + 2\delta}{\theta + 2} \]

\[ = \frac{\theta + 2\delta}{\theta + 2} \left( \frac{1 - \delta}{\theta + 8} + \frac{\theta + 4\delta}{\theta + 8} \frac{1}{V} \right) \]