Inference of SIG
-- Inside Outside Algorithm --

Song-Chun Zhu, Sinisa Todorovic, and Ales Leonardis

At CVPR, Providence, Rhode Island
June 16, 2012
Inside-Outside Algorithm

• Notation:
  
  – Let $O = O_1, O_2, ..., O_T$ be the observation sequence generated by a SCFG $G$.
  
  – let $i, j, k$ be integer numbers corresponding to each of the non-terminal symbols.
  
  – Let $m$ be an integer corresponding to a terminal symbol.
  
  – The grammar $G$ has the Chomsky Normal Form

  $i \rightarrow jk$ \hspace{1cm} $i \rightarrow m$
Inside-Outside Algorithm

• Notation:
  – Probability matrix $A$ and $B$
    
    $a[i, j, k] = P(i \rightarrow jk)$
    $b[i, m] = P(i \rightarrow m)$
  – $a[i, j, k]$ is the probability that the non-terminal symbol $i$ generate the pair of non-terminal symbols $j$ and $k$
  – $b[i, m]$ is the probability that the non-terminal symbol $i$ generate a single terminal symbol $m$
Inner Probability

e(s, t, i) = Probability of the non-terminal symbol $i$ generating the observation $O(s),...,O(t)$

calculation of inner probabilities
Computation of Inner Probability

• When $s = t$
  
  $$e(s, s, i) = P(i \rightarrow O(s)) = b[i, O(s)]$$

• When $s \neq t$
  
  $$e(s, t, i) = \sum_{j,k} \sum_{r=s}^{t-1} a[i, j, k]e(s, r, j)e(r + 1, t, k)$$

• The quantity $e$ can be computed recursively by determining $e$ for all sequences of length 1, then all sequences of length 2, and so on.
Outer Probability

\[ f \]

\[ I \]

\[ 0(1)...0(s-1) \quad 0(s)...0(t) \quad 0(t+1)...0(T) \]
Outer Probability

• Define outer probability $f$ as

$$f(s, t, i) = P(S \Rightarrow O(1)...O(s-1), i, O(t+1)...O(T))$$

• $f$ can be computed by

$$f(s, t, i) = \sum_{j, k} \left[ \sum_{r=1}^{s-1} f(r, t, j)a[j, k, i]e(r, s-1, k) \right]$$

$$+ \sum_{r=t+1}^{T} f(s, r, j)a[j, i, k]e(t+1, r, k)$$

• And

$$f(1,T,i) = \begin{cases} 
1 & \text{if } (i = S) \\
0 & \text{otherwise} 
\end{cases}$$
Computation of Outer Probability

Because a non-terminal node allows binary splits

\[ i \rightarrow j \ k \]
Inside-Outside Algorithm: Example

<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP VP</td>
<td>1.0</td>
</tr>
<tr>
<td>PP → P NP</td>
<td>1.0</td>
</tr>
<tr>
<td>VP → V NP</td>
<td>0.7</td>
</tr>
<tr>
<td>VP → VP PP</td>
<td>0.3</td>
</tr>
<tr>
<td>P → with</td>
<td>1.0</td>
</tr>
<tr>
<td>V → saw</td>
<td>1.0</td>
</tr>
<tr>
<td>NP → NP PP</td>
<td>0.4</td>
</tr>
<tr>
<td>NP → astronomers</td>
<td>0.1</td>
</tr>
<tr>
<td>NP → ears</td>
<td>0.18</td>
</tr>
<tr>
<td>NP → saw</td>
<td>0.04</td>
</tr>
<tr>
<td>NP → stars</td>
<td>0.18</td>
</tr>
<tr>
<td>NP → telescopes</td>
<td>0.1</td>
</tr>
</tbody>
</table>

\[ x = \text{astronomers saw stars with ears} \]
Inside-Outside Algorithm: Example

t_1:
\[
\begin{array}{c}
S_{1.0} \\
\mid \\
NP_{0.1} \\
\mid \\
astronomers \\
\mid \\
V_{1.0} \\
\mid \\
saw \\
\mid \\
NP_{0.4} \\
\mid \\
NP_{0.18} \\
\mid \\
stars \\
\mid \\
P_{1.0} \\
\mid \\
PP_{1.0} \\
\mid \\
with \\
\mid \\
NP_{0.18} \\
\mid \\
ears
\end{array}
\]

t_2:
\[
\begin{array}{c}
S_{1.0} \\
\mid \\
NP_{0.1} \\
\mid \\
astronomers \\
\mid \\
V_{1.0} \\
\mid \\
saw \\
\mid \\
NP_{0.18} \\
\mid \\
stars \\
\mid \\
P_{1.0} \\
\mid \\
PP_{1.0} \\
\mid \\
with \\
\mid \\
NP_{0.18} \\
\mid \\
ears
\end{array}
\]
Inside-Outside Algorithm: Example

- $e(1,1,S)=0 \quad e(1,1,PP)=0 \quad e(1,1,VP)=0 \quad e(1,1,NP)=0.1 \quad e(1,1,V)=0 \quad e(1,1,P)=0$
- $e(2,2,S)=0 \quad e(2,2,PP)=0 \quad e(2,2,VP)=0 \quad e(2,2,NP)=0.04 \quad e(2,2,V)=1 \quad e(2,2,P)=0$
- $e(3,3,S)=0 \quad e(3,3,PP)=0 \quad e(3,3,VP)=0 \quad e(3,3,NP)=0.18 \quad e(3,3,V)=0 \quad e(3,3,P)=0$
- $e(4,4,S)=0 \quad e(4,4,PP)=0 \quad e(4,4,VP)=0 \quad e(4,4,NP)=0 \quad e(4,4,V)=0 \quad e(4,4,P)=1$
- $e(5,5,S)=0 \quad e(5,5,PP)=0 \quad e(5,5,VP)=0 \quad e(5,5,NP)=0.18 \quad e(5,5,V)=0 \quad e(5,5,P)=0$

- $e(1,2,S)=0 \quad e(1,2,PP)=0 \quad e(1,2,VP)=0 \quad e(1,2,NP)=0 \quad e(1,2,V)=0 \quad e(1,2,P)=0$
- $e(2,3,S)=0 \quad e(2,3,PP)=0 \quad e(2,3,VP)=0 \quad e(2,3,NP)=0 \quad e(2,3,V)=0 \quad e(2,3,P)=0$

- ....
- $f(1,5,S)=1 \quad f(2,5,VP)=f(1,5,S) \times a[S,NP,VP] \times e(1,1,NP)=0.1$
- $f(3,5,NP)=f(2,5,VP) \times a[VP,V,NNP] \times e(2,2,V)=0.07$

- $P(O|G) = 0.015876$