Jurafsky’s Probabilistic Model of Syntactic Disambiguation

Frank Keller

Computerlinguistik
Universität des Saarlandes

keller@coli.uni-sb.de
Overview

Jurafsky’s (1996) approach:

- probabilistic model of lexical and syntactic access and disambiguation;
- accounts for psycholinguistic data using concepts from computational linguistics: probabilistic CFGs, Bayesian modeling frame probabilities;
- focus here: syntactic disambiguation in human sentence processing.

Overview of the lecture:

- data to be modeled: frame preferences, garden paths;
- architecture: serial, parallel, limited parallel;
- probabilistic CFGs, frame probabilities;
- examples for frame preferences, garden paths;
- comparison with other models; problems and issues.
Frame Preferences

(1) The women discussed the dogs on the beach.
   a. The women discussed the dogs which were on the beach. (90%)
   b. The women discussed them (the dogs) while on the beach. (10%)

(2) The women kept the dogs on the beach.
   a. The women kept the dogs which were on the beach. (5%)
   b. The women discussed them (the dogs) while on the beach. (95%)
Garden Paths

Main Clause vs. Reduced Relative Ambiguity

(3) # The horse raced past the barn fell.
(4) # The teachers taught by the Berlitz method passed the test.
(5) ? The children taught by the Berlitz method passed the test.

Lexical Category Ambiguity

(6) # The complex houses married and single students and their families.
(7) # The warehouse fires destroyed all the buildings.
(8) # The warehouse fires a dozen employees each year.
(9) # The prime number few.
(10) # The old man the boats.
(11) # The grappling hooks on to the enemy ship.

Frame Ambiguity

(12) # The landlord painted all the walls with cracks.
(13) # Ross baked the cake in the freezer.
Parser Architectures

Serial Parser

- build parse trees through successive rule selection;
- if more than one rule applies (choice point), choose one possible tree based on a selection rule;
- if the tree turns out to be impossible, return to the choice point (backtracking) and reparse from there;
- example for selection rule: minimal attachment (choose the tree with the least nodes).

Parallel Parser

- build parse trees through successive rule selection;
- if more than one rule applies, create a new tree for each rule;
- pursue all possibilities in parallel;
- if one turns out to be impossible, drop it;
- problem: number of parse trees can grow exponentially.
- solution: bounded parallelism, only pursue a limited number of possibilities (prune trees).
Modeling Human Parsing

Serial Parser

- garden path means: wrong tree was selected at a choice point;
- backtracking occurs, causes increased processing times.

Parallel Parser

- garden path means: correct tree was pruned;
- backtracking occurs, causes increased processing times.

Jurafsky (1996) assumes bounded parallelism in a parsing model based on probabilistic CFGs.

Pruning occurs if a parse tree is sufficiently improbably (beam search algorithm).
Probabilistic Context-free Grammars

- context-free rules annotated with probabilities;
- probabilities of all rules with the same lefthand side sum to one;
- probability of a parse is the product of the probabilities of all rules applied in the parse.

Example (Manning and Schütze 1999)

<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>
Probabilistic Context-free Grammars

Example (Manning and Schütze 1999)

\[ t_1: \]

\[ P(t_1) = 1.0 \times 0.1 \times 0.7 \times 1.0 \times 0.4 \times 0.18 \times 1.0 \times 1.0 \times 0.18 = 0.0009072 \]
Probabilistic Context-free Grammars

Example (Manning and Schütze 1999)

t_2:

\[
P(t_1) = 1.0 \times 0.1 \times 0.3 \times 0.7 \times 1.0 \times 0.18 \times 1.0 \times 1.0 \times 0.18 = 0.0006804
\]
**Frame Probabilities**

Complements of *keep*:

<table>
<thead>
<tr>
<th>Frame Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>keep the prices reasonable</td>
</tr>
<tr>
<td>VP</td>
<td>keep his foes guessing</td>
</tr>
<tr>
<td>VP</td>
<td>keep their eyes peeled</td>
</tr>
<tr>
<td>PRT</td>
<td>keep the people in</td>
</tr>
<tr>
<td>PP</td>
<td>keep his nerves from jangling</td>
</tr>
</tbody>
</table>

Frame probabilities computed from the Penn Treebank:

<table>
<thead>
<tr>
<th>Frame</th>
<th>Syntax</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>discuss</td>
<td>⟨NP PP⟩</td>
<td>.24</td>
</tr>
<tr>
<td></td>
<td>⟨NP⟩</td>
<td>.76</td>
</tr>
<tr>
<td>keep</td>
<td>⟨NP XP[pred +]⟩</td>
<td>.81</td>
</tr>
<tr>
<td></td>
<td>⟨NP⟩</td>
<td>.19</td>
</tr>
</tbody>
</table>
Modeling Frame Preferences

\[ p(\text{keep}, \langle \text{NP XP[pred +]} \rangle) = 0.81 \]

\[ \text{VP} \rightarrow \text{V NP XP} \quad 0.15 \]

\[ t_1 : \]

```
   VP
   /  |
  V   NP   PP
 |    /    |
keep the dogs on the beach
```

\[ p(t_1) = 0.15 \times 0.81 = 0.12 \text{ (preferred)} \]
**Modeling Frame Preferences**

\[ p(\text{keep, } \langle \text{NP} \rangle) = 0.19 \]

\[ \text{VP} \rightarrow \text{V NP} \quad 0.39 \]
\[ \text{NP} \rightarrow \text{NP XP} \quad 0.14 \]

\( t_2: \)

```
  VP
   /\  \
  V   NP
    /\    /
   keep NP PP
     /\  |
    the dogs on the beach
```

\[ p(t_2) = 0.19 \times 0.39 \times 0.14 = 0.01 \quad \text{(dispreferred)} \]
Modeling Frame Preferences

\[ p(\text{discuss, } \langle \text{NP PP} \rangle) = 0.24 \]

\[ \text{VP} \rightarrow \text{V NP XP} \quad 0.15 \]

\[ t_1: \]

```
VP
|---- V
| discuss
|---- NP
| the dogs
|---- PP
| on the beach
```

\[ p(t_1) = 0.15 \times 0.24 = 0.036 \text{ (dispreferred)} \]
**Modeling Frame Preferences**

\[ p(\text{discuss}, \langle \text{NP} \rangle) = 0.76 \]

\[ \text{VP} \rightarrow \text{V NP} \quad 0.39 \]

\[ \text{NP} \rightarrow \text{NP XP} \quad 0.14 \]

\( t_2 : \)

\[
\begin{align*}
\text{VP} \\
&\text{V} \\
&\text{discuss} \\
&\text{NP} \\
&\text{the dogs} \\
&\text{PP} \\
&\text{on the beach}
\end{align*}
\]

\[ p(t_2) = 0.76 \times 0.39 \times 0.14 = 0.041 \text{ (preferred)} \]
**Modeling Garden Path Effects**

Garden path caused by construction probabilities

\[
\begin{align*}
S & \rightarrow \text{NP} \ldots \quad 0.92 \\
\text{NP} & \rightarrow \text{Det} \ \text{Adj} \ \text{N} \quad 0.28 \\
\text{N} & \rightarrow \text{ROOT} \ s \quad 0.23 \\
\text{N} & \rightarrow \text{house} \quad 0.0024 \\
\text{Adj} & \rightarrow \text{complex} \quad 0.00086
\end{align*}
\]

\( t_1: \)

\[
\begin{array}{c}
S \\
\downarrow \\
\text{NP} \\
\downarrow \\
\text{Det} & \text{Adj} & \text{N} \\
\downarrow & \downarrow & \downarrow \\
\text{the} & \text{complex} & \text{houses}
\end{array}
\]

\[
p(t_1) = 1.2 \times 10^{-7} \text{ (preferred)}
\]
Modeling Garden Path Effects

NP $\rightarrow$ Det N 0.63
S $\rightarrow$ [NP VP]V ... 0.48
N $\rightarrow$ complex 0.000029
V $\rightarrow$ house 0.0006
V $\rightarrow$ ROOT s 0.086

$t_1$:

```
S
  /   \
/     \   
NP     VP
 |      /  \
the   N    V
|  |    /   \
complex    houses
```

$p(t_1) = 4.5 \times 10^{-10}$ (dispreferred)
Modeling Garden Path Effects

\[ S \rightarrow \text{NP} \ldots \quad 0.92 \]
\[ \text{NP} \rightarrow \text{Det N N} \quad 0.28 \]
\[ \text{N} \rightarrow \text{fire} \quad 0.00072 \]
\[ \text{N} \rightarrow \text{ROOT s} \quad 0.23 \]

\[ t_1: \]

\[
\begin{array}{c}
S \\
\quad \text{NP} \\
\quad \quad \text{Det} \quad \text{N} \quad \text{N} \\
\quad \quad \quad \text{the} \quad \text{warehouse} \quad \text{fires}
\end{array}
\]

\[ p(t_1) = 4.2 \times 10^{-5} \text{ (preferred)} \]
Modeling Garden Path Effects

NP → Det N 0.63
S → [NP VP[V ... 0.48
V → fire 0.00042
V → ROOT s 0.086

\[ t_1: \]

\[
\begin{array}{c}
S \\
NP & VP \\
Det & N & V \\
the & warehouse & fires
\end{array}
\]

\[ p(t_1) = 1.1 \times 10^{-5} \text{ (dispreferred)} \]
Modeling Garden Path Effects

Garden path caused by construction probabilities and frame probabilities

\[ p(\text{race}, \langle \text{NP} \rangle) = 0.92 \]

\[ t_1: \]

\[
S \\
NP \quad VP \\
\text{the horse} \quad \text{raced}
\]

\[ p(t_1) = 0.92 \text{ (preferred)} \]
Modeling Garden Path Effects

\[ p(\text{race}, \langle \text{NP NP} \rangle) = 0.08 \]

\[ \text{NP} \rightarrow \text{NP XP} \quad 0.14 \]

\( t_2: \)

\[
\begin{array}{c}
S \\
NP \\
\text{the horse} \quad \text{raced} \\
\end{array}
\]

\[ p(t_1) = 0.0112 \text{ (dispreferred)} \]
Modeling Garden Path Effects

\[ p(\text{find}, \langle \text{NP} \rangle) = 0.38 \]

\( t_1: \)

\[ S \]
\[ \text{NP} \quad \text{VP} \]
\[ \text{the bird} \quad \text{found} \]

\[ p(t_1) = 0.38 \text{ (preferred)} \]
Modeling Garden Path Effects

\[ p(\text{find}, \{\text{NP NP}\}) = 0.62 \]

NP \rightarrow NP \; XP \; 0.14

\( t_2: \)

\[
\begin{array}{c}
S \\
\text{NP} \\
\text{NP} \quad \text{VP} \\
\text{the bird} \quad \text{found}
\end{array}
\]

\[ p(t_1) = 0.0868 \; \text{(dispreferred)} \]
Setting the Beam Width

Crucial assumption: if the relative probability of a tree falls below a certain value, then it will be pruned.

<table>
<thead>
<tr>
<th>sentence</th>
<th>probability ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>the complex houses ...</td>
<td>267:1</td>
</tr>
<tr>
<td>the horse raced ...</td>
<td>82:1</td>
</tr>
<tr>
<td>the warehouse fires ...</td>
<td>3.8:1</td>
</tr>
<tr>
<td>the bird found ...</td>
<td>3.7:1</td>
</tr>
</tbody>
</table>

Assumption: a garden path occurs if the probability ratio is higher than 5:1.
Open Issues

- Incrementality: Can we make more fine-grained predictions of the time course of ambiguity resolution?
- Coverage: Jurafsky used hand-crafted examples. Can we use a probabilistic parser that is trained on a real corpus?
- Memory Limitations: How can we augment the model to take memory limitations into account (e.g., center embedding)?
- Crosslinguistics: does this model work for languages other than English?
References
