Admin

Assignment 3 out today: due next Monday

Quiz

Context free grammar

\[ S \rightarrow \text{NP} \quad \text{VP} \]

left hand side  
right hand side
(single symbol)  (one or more symbols)

Formally...

\[ G = (\text{NT}, T, P, S) \]

NT: finite set of nonterminal symbols

T: finite set of terminal symbols, NT and T are disjoint

P: finite set of productions of the form
\[ A \rightarrow \alpha, \ A \in \text{NT} \text{ and } \alpha \in (T \cup \text{NT})^* \]

S \in \text{NT}: start symbol
Many possible CFGs for English, here is an example (fragment):

```plaintext
CFG: Example

S → NP VP
VP → V NP
NP → DetP N | DetP AdjP N
AdjP → Adj | Adv AdjP
N → kid | dog
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the
```

Derivations in a CFG

```plaintext
Derivations in a CFG

S → NP VP
VP → V NP
NP → DetP N | DetP AdjP N
AdjP → Adj | Adv AdjP
N → kid | dog
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the
```

What can we do?

```plaintext
What can we do?

S → NP VP
VP → V NP
NP → DetP N | DetP AdjP N
AdjP → Adj | Adv AdjP
N → kid | dog
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the
```

```plaintext
What can we do?

S → NP VP
VP → V NP
NP → DetP N | DetP AdjP N
AdjP → Adj | Adv AdjP
N → kid | dog
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the
```

```plaintext
What can we do?

S → NP VP
VP → V NP
NP → DetP N | DetP AdjP N
AdjP → Adj | Adv AdjP
N → kid | dog
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the
```

```plaintext
What can we do?

S → NP VP
VP → V NP
NP → DetP N | DetP AdjP N
AdjP → Adj | Adv AdjP
N → kid | dog
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the
```
Derivations in a CFG

\[
\begin{align*}
S & \rightarrow NP \ VP \\
VP & \rightarrow V \ NP \\
NP & \rightarrow DetP \ N \ | \ DetP \ AdjP \ N \\
AdjP & \rightarrow Adj \ | \ Adv \ AdjP \\
N & \rightarrow \text{kid} \ | \ \text{dog} \\
V & \rightarrow \text{sees} \ | \ \text{likes} \\
Adj & \rightarrow \text{big} \ | \ \text{small} \\
Adv & \rightarrow \text{very} \\
DetP & \rightarrow a \ | \ the
\end{align*}
\]
Derivations in a CFG

$S \rightarrow NP \ VP$
$VP \rightarrow V \ NP$
$NP \rightarrow DetP \ N \mid DetP \ AdjP \ N$
$AdjP \rightarrow Adj \mid Adv \ AdjP$
$N \rightarrow \text{kid} \mid \text{dog}$
$V \rightarrow \text{sees} \mid \text{likes}$
$Adj \rightarrow \text{big} \mid \text{small}$
$Adv \rightarrow \text{very}$
$DetP \rightarrow \text{a} \mid \text{the}$

Derivations of CFGs

String rewriting system: we derive a string

Derivation history shows the constituent tree:

```
the kid likes a dog
```
Parsing

Parsing is the field of NLP interested in automatically determining the syntactic structure of a sentence.

Parsing can be thought of as determining what sentences are "valid" English sentences.

As a byproduct, we often can get the structure.

Given a CFG and a sentence, determine the possible parse tree(s).

I eat sushi with tuna

What parse trees are possible for this sentence?

How did you do it?

What if the grammar is much larger?

What is the difference between these parses?

How can we decide between these?
A Simple PCFG

Probabilities!

- $S \rightarrow NP\ VP$ with probability 1.0
- $NP \rightarrow NP\ PP$ with probability 0.4
- $VP \rightarrow V\ NP$ with probability 0.7
- $VP \rightarrow VP\ PP$ with probability 0.3
- $PP \rightarrow P\ NP$ with probability 1.0
- $P \rightarrow with$ with probability 1.0
- $V \rightarrow saw$ with probability 1.0
- $NP \rightarrow astronomers$ with probability 0.1
- $NP \rightarrow ears$ with probability 0.18
- $NP \rightarrow stars$ with probability 0.18
- $NP \rightarrow telescope$ with probability 0.1

Just like n-gram language modeling, PCFGs break the sentence generation process into smaller steps/probabilities.

The probability of a parse is the product of the PCFG rules.

What are the different interpretations here?
Which do you think is more likely?
Parsing problems

Pick a model
- e.g. CFG, PCFG, ...

Train (or learn) a model
- What CFG/PCFG rules should I use?
- Parameters (e.g. PCFG probabilities)?
- What kind of data do we have?

Parsing
- Determine the parse tree(s) given a sentence

PCFG: Training

If we have example parsed sentences, how can we learn a set of PCFGs?

Tree Bank

Supervised PCFG Training

S → NP VP 0.9
S → NP PP 0.1
NP → Det N 0.6
NP → NP PP 0.3
NP → Prop N 0.2
V → saw 1.0
V → V NP 0.7
V → V VP 0.3
V → V PP 0.3

English

Extracting the rules

What CFG rules occur in this tree?

I eat sushi with tuna

S → NP VP
NP → PRP
PRP → I
VP → V NP
V → eat
NP → N PP
N → sushi
PP → IN N
IN → with
N → tuna

Estimating PCFG Probabilities

We can extract the rules from the trees

S → NP VP
NP → PRP
PRP → I
VP → V NP
V → eat
NP → N PP
N → sushi
...
Estimating PCFG Probabilities

- Extract the rules from the trees
- Calculate the probabilities using MLE

\[ P(\alpha \rightarrow \beta | \alpha) = \frac{\text{count}(\alpha \rightarrow \beta)}{\text{count}(\alpha)} \]

Grammar Equivalence

- Weak equivalence: grammars generate the same set of strings
  - Grammar 1: \( NP \rightarrow \text{DetP N} \) and \( \text{DetP} \rightarrow \alpha \mid \text{the} \)
  - Grammar 2: \( NP \rightarrow \alpha N \mid \text{the} N \)
- Strong equivalence: grammars have the same set of derivation trees
  - With CFGs, possible only with useless rules
  - Grammar 2: \( NP \rightarrow \alpha N \mid \text{the} N \)
  - Grammar 3: \( NP \rightarrow \alpha N \mid \text{the} N, \text{DetP} \rightarrow \text{many} \)

Estimating PCFG Probabilities

<table>
<thead>
<tr>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>S \rightarrow NP VP</td>
</tr>
<tr>
<td>S \rightarrow V NP</td>
</tr>
<tr>
<td>S \rightarrow VP PP</td>
</tr>
<tr>
<td>NP \rightarrow N</td>
</tr>
<tr>
<td>NP \rightarrow N PP</td>
</tr>
<tr>
<td>NP \rightarrow DT N</td>
</tr>
</tbody>
</table>

\[ P( S \rightarrow V NP) = \frac{\text{count}(S \rightarrow V NP)}{\text{count}(S)} = \frac{3}{15} \]
Normal Forms

There are weakly equivalent normal forms (Chomsky Normal Form, Greibach Normal Form)

A CFG is in Chomsky Normal Form (CNF) if all productions are of one of two forms:
- $A \rightarrow BC$ with $A, B, C$ nonterminals
- $A \rightarrow a$, with $A$ a nonterminal and $a$ a terminal

Every CFG has a weakly equivalent CFG in CNF

Probabilistic Grammar Conversion

<table>
<thead>
<tr>
<th>Original Grammar</th>
<th>Chomsky Normal Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S \rightarrow NP VP$</td>
<td>0.8</td>
</tr>
<tr>
<td>$S \rightarrow Aux NP VP$</td>
<td>0.1</td>
</tr>
<tr>
<td>$S \rightarrow VP$</td>
<td>0.1</td>
</tr>
<tr>
<td>$NP \rightarrow Pronoun$</td>
<td>0.2</td>
</tr>
<tr>
<td>$NP \rightarrow Proper Noun$</td>
<td>0.2</td>
</tr>
<tr>
<td>$NP \rightarrow Det Nominal$</td>
<td>0.4</td>
</tr>
<tr>
<td>$Nominal \rightarrow Noun$</td>
<td>0.3</td>
</tr>
<tr>
<td>$Nominal \rightarrow Nominal Noun$</td>
<td>0.2</td>
</tr>
<tr>
<td>$Nominal \rightarrow Nominal PP$</td>
<td>0.5</td>
</tr>
<tr>
<td>$VP \rightarrow Verb$</td>
<td>0.2</td>
</tr>
<tr>
<td>$VP \rightarrow Verb NP$</td>
<td>0.5</td>
</tr>
<tr>
<td>$VP \rightarrow VP PP$</td>
<td>0.3</td>
</tr>
<tr>
<td>$PP \rightarrow Prep NP$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Grammar questions

Can we determine if a sentence is grammatical?

Given a sentence, can we determine the syntactic structure?

Can we determine how likely a sentence is to be grammatical? to be an English sentence?

Can we generate candidate, grammatical sentences?