

CS 6501 Natural Language Processing

Probabilistic Context-Free Grammars

Yangfeng Ji

September 24, 2018

Department of Computer Science
University of Virginia



ENGINEERING

1. Probabilistic CFGs
2. Probabilistic CKY Algorithm

Based on slides from [Collins, 2017, Smith, 2017]

Probabilistic CFGs

A Probabilistic Context-Free Grammar (PCFG)

- ▶ N : a set of non-terminal symbols
- ▶ $S \in N$: a distinguished start symbol
- ▶ Σ : a set of terminal symbols

S	\Rightarrow	NP	VP	1.0
VP	\Rightarrow	Vi		0.4
VP	\Rightarrow	Vt	NP	0.4
VP	\Rightarrow	VP	PP	0.2
NP	\Rightarrow	DT	NN	0.3
NP	\Rightarrow	NP	PP	0.7
PP	\Rightarrow	P	NP	1.0

Vi	\Rightarrow	sleeps	1.0
Vt	\Rightarrow	saw	1.0
NN	\Rightarrow	man	0.7
NN	\Rightarrow	woman	0.2
NN	\Rightarrow	telescope	0.1
DT	\Rightarrow	the	1.0
IN	\Rightarrow	with	0.5
IN	\Rightarrow	in	0.5

Probability of a Tree

The probability of a tree t with rules $\{\alpha_i \rightarrow \beta_i\}$, such as

$S \rightarrow \text{NP VP}, \text{NP} \rightarrow \text{DT NN}, \dots, \text{Vi} \rightarrow \text{sleeps}$

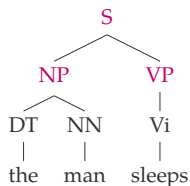
is

$$\begin{aligned} P(t) &= \underbrace{\prod_{i=1}^n P(\alpha_i \rightarrow \beta_i)}_{\text{production rule form}} \\ &= \underbrace{\prod_{i=1}^n P(\beta_i \mid \alpha_i)}_{\text{Standard conditional prob form}} \end{aligned} \tag{1}$$

An Example

S	⇒	NP	VP	1.0
VP	⇒	Vi		0.4
VP	⇒	Vt	NP	0.4
VP	⇒	VP	PP	0.2
NP	⇒	DT	NN	0.3
NP	⇒	NP	PP	0.7
PP	⇒	P	NP	1.0

Vi	⇒	sleeps	1.0
Vt	⇒	saw	1.0
NN	⇒	man	0.7
NN	⇒	woman	0.2
NN	⇒	telescope	0.1
DT	⇒	the	1.0
IN	⇒	with	0.5
IN	⇒	in	0.5



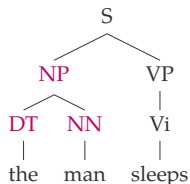
$$P(t) = P(\text{NP VP} \mid S)$$

(2)

An Example

S	⇒	NP	VP	1.0
VP	⇒	Vi		0.4
VP	⇒	Vt	NP	0.4
VP	⇒	VP	PP	0.2
NP	⇒	DT	NN	0.3
NP	⇒	NP	PP	0.7
PP	⇒	P	NP	1.0

Vi	⇒	sleeps	1.0
Vt	⇒	saw	1.0
NN	⇒	man	0.7
NN	⇒	woman	0.2
NN	⇒	telescope	0.1
DT	⇒	the	1.0
IN	⇒	with	0.5
IN	⇒	in	0.5



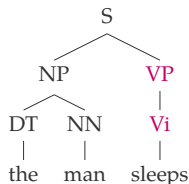
$$P(t) = P(\text{NP VP} \mid S) \cdot P(\text{DT NN} \mid \text{NP})$$

(2)

An Example

S	⇒	NP	VP	1.0
VP	⇒	Vi		0.4
VP	⇒	Vt	NP	0.4
VP	⇒	VP	PP	0.2
NP	⇒	DT	NN	0.3
NP	⇒	NP	PP	0.7
PP	⇒	P	NP	1.0

Vi	⇒	sleeps	1.0
Vt	⇒	saw	1.0
NN	⇒	man	0.7
NN	⇒	woman	0.2
NN	⇒	telescope	0.1
DT	⇒	the	1.0
IN	⇒	with	0.5
IN	⇒	in	0.5



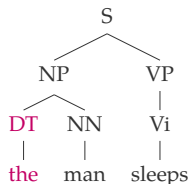
$$P(t) = P(\text{NP VP} \mid S) \cdot P(\text{DT NN} \mid \text{NP}) \cdot P(\text{Vi} \mid \text{VP})$$

(2)

An Example

S	⇒	NP	VP	1.0
VP	⇒	Vi		0.4
VP	⇒	Vt	NP	0.4
VP	⇒	VP	PP	0.2
NP	⇒	DT	NN	0.3
NP	⇒	NP	PP	0.7
PP	⇒	P	NP	1.0

Vi	⇒	sleeps	1.0
Vt	⇒	saw	1.0
NN	⇒	man	0.7
NN	⇒	woman	0.2
NN	⇒	telescope	0.1
DT	⇒	the	1.0
IN	⇒	with	0.5
IN	⇒	in	0.5

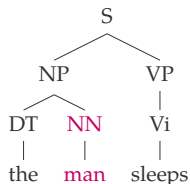


$$P(t) = P(\text{NP VP} \mid S) \cdot P(\text{DT NN} \mid \text{NP}) \cdot P(\text{Vi} \mid \text{VP}) \\ \cdot P(\text{the} \mid \text{DT}) \quad (2)$$

An Example

S	⇒	NP	VP	1.0
VP	⇒	Vi		0.4
VP	⇒	Vt	NP	0.4
VP	⇒	VP	PP	0.2
NP	⇒	DT	NN	0.3
NP	⇒	NP	PP	0.7
PP	⇒	P	NP	1.0

Vi	⇒	sleeps	1.0
Vt	⇒	saw	1.0
NN	⇒	man	0.7
NN	⇒	woman	0.2
NN	⇒	telescope	0.1
DT	⇒	the	1.0
IN	⇒	with	0.5
IN	⇒	in	0.5

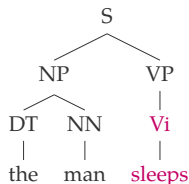


$$\begin{aligned} P(t) = & P(\text{NP VP} \mid S) \cdot P(\text{DT NN} \mid \text{NP}) \cdot P(\text{Vi} \mid \text{VP}) \\ & \cdot P(\text{the} \mid \text{DT}) \cdot P(\text{man} \mid \text{NN}) \end{aligned} \quad (2)$$

An Example

S	⇒	NP	VP	1.0
VP	⇒	Vi		0.4
VP	⇒	Vt	NP	0.4
VP	⇒	VP	PP	0.2
NP	⇒	DT	NN	0.3
NP	⇒	NP	PP	0.7
PP	⇒	P	NP	1.0

Vi	⇒	sleeps	1.0
Vt	⇒	saw	1.0
NN	⇒	man	0.7
NN	⇒	woman	0.2
NN	⇒	telescope	0.1
DT	⇒	the	1.0
IN	⇒	with	0.5
IN	⇒	in	0.5



$$\begin{aligned} P(t) = & P(\text{NP VP} \mid S) \cdot P(\text{DT NN} \mid \text{NP}) \cdot P(\text{Vi} \mid \text{VP}) \\ & \cdot P(\text{the} \mid \text{DT}) \cdot P(\text{man} \mid \text{NN}) \cdot P(\text{sleeps} \mid \text{Vi}) \end{aligned} \quad (2)$$

Properties of PCFGs

- ▶ Assigns a **probability** to each derivation, or parse-tree, allowed by the underlying CFG

Properties of PCFGs

- ▶ Assigns a **probability** to each derivation, or parse-tree, allowed by the underlying CFG
- ▶ If one sentence has **more than one derivations**, we can rank them based on their probabilities

Properties of PCFGs

- ▶ Assigns a **probability** to each derivation, or parse-tree, allowed by the underlying CFG
- ▶ If one sentence has **more than one derivations**, we can rank them based on their probabilities
- ▶ The **most likely parse tree** for a sentence is

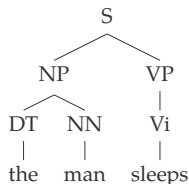
$$\arg \max_{t \in \mathcal{T}(s)} P(t|s) \quad (3)$$

where $\mathcal{T}(s)$ is the set of all possible parse trees of sentence s .

Probabilistic CKY Algorithm

Score of Parse Trees: An example

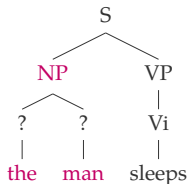
$$P(\mathbf{t} \mid \mathbf{s}) = P(\text{NP VP} \mid \text{S}) \cdot P(\text{DT NN} \mid \text{NP}) \cdot P(\text{Vi} \mid \text{VP}) \\ \cdot P(\text{the} \mid \text{DT}) \cdot P(\text{man} \mid \text{NN}) \cdot P(\text{sleeps} \mid \text{Vi}) \quad (4)$$



- ▶ Decoding $\arg \max_{\mathbf{t}} P(\mathbf{t} \mid \mathbf{s})$
- ▶ Effect of change on non-terminal node
- ▶ Similar phenomenon is handled by Viterbi decoding in HMM and CRF

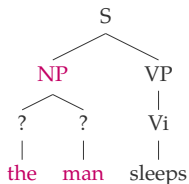
Notations (I)

- ▶ X : non-terminal node
- ▶ i, j : word indices, $1 \leq i \leq j \leq n$
- ▶ $\mathcal{T}(i, j, X)$: the set of all parse trees for words x_i, \dots, x_j with X as the root
- ▶ Example: $\mathcal{T}(1, 2, \text{NP})$



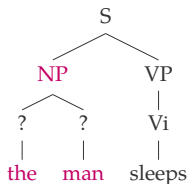
Notations (II)

- ▶ X : non-terminal node
- ▶ i, j : word indices, $1 \leq i \leq j \leq n$
- ▶ $\pi(i, j, X) = \max_{t \in \mathcal{T}(i, j, X)} P(t)$
- ▶ Example: $\pi(1, 2, \text{NP})$



Notations (II)

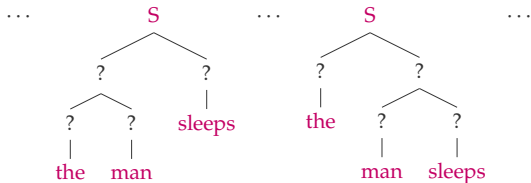
- ▶ X : non-terminal node
- ▶ i, j : word indices, $1 \leq i \leq j \leq n$
- ▶ $\pi(i, j, X) = \max_{t \in \mathcal{T}(i, j, X)} P(t)$
- ▶ Example: $\pi(1, 2, \text{NP})$



- ▶ $\pi(i, j, X) = 0$, if $\mathcal{T}(i, j, X) = \emptyset$, for example $\pi(1, 2, \text{VP})$

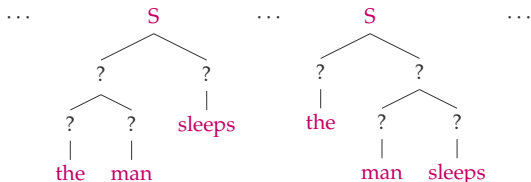
Special Cases of $\mathcal{T}(i, j, X)$ and $\pi(i, j, X)$

► $\mathcal{T}(1, n, S)$

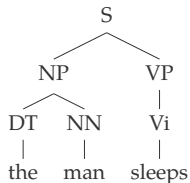


Special Cases of $\mathcal{T}(i, j, X)$ and $\pi(i, j, X)$

► $\mathcal{T}(1, n, S)$



► $\pi(1, n, S)$: the score of the optimal tree



Special Cases of $\mathcal{T}(i, j, X)$ and $\pi(i, j, X)$

For the example sentence the man sleeps

- ▶ $\mathcal{T}(1, 1, X)$ if $X = \text{DT}$

DT
|
the

- ▶ $\pi(1, 1, \text{DT}) = 1$

S	⇒	NP	VP	1.0
VP	⇒	Vi		0.4
VP	⇒	Vt	NP	0.4
VP	⇒	VP	PP	0.2
NP	⇒	DT	NN	0.3
NP	⇒	NP	PP	0.7
PP	⇒	P	NP	1.0

Vi	⇒	sleeps	1.0
Vt	⇒	saw	1.0
NN	⇒	man	0.7
NN	⇒	woman	0.2
NN	⇒	telescope	0.1
DT	⇒	the	1.0
IN	⇒	with	0.5
IN	⇒	in	0.5

Special Cases of $\mathcal{T}(i, j, X)$ and $\pi(i, j, X)$

For the example sentence the man sleeps

- ▶ $\mathcal{T}(1, 1, X)$ if $X = \text{DT}$

DT
|
the

- ▶ $\pi(1, 1, \text{DT}) = 1$
- ▶ What if $X = \text{NN}$?

S	⇒	NP	VP	1.0
VP	⇒	Vi		0.4
VP	⇒	Vt	NP	0.4
VP	⇒	VP	PP	0.2
NP	⇒	DT	NN	0.3
NP	⇒	NP	PP	0.7
PP	⇒	P	NP	1.0

Vi	⇒	sleeps	1.0
Vt	⇒	saw	1.0
NN	⇒	man	0.7
NN	⇒	woman	0.2
NN	⇒	telescope	0.1
DT	⇒	the	1.0
IN	⇒	with	0.5
IN	⇒	in	0.5

Special Cases of $\mathcal{T}(i, j, X)$ and $\pi(i, j, X)$

For the example sentence the man sleeps

- ▶ $\mathcal{T}(1, 1, X)$ if $X = \text{DT}$

DT
|
the

- ▶ $\pi(1, 1, \text{DT}) = 1$
- ▶ What if $X = \text{NN}$?

$$\mathcal{T}(1, 1, \text{NN}) = \emptyset$$

$$\pi(1, 1, \text{NN}) = 0$$

because there is no such rule
 $\text{NN} \rightarrow \text{the}$

S	\Rightarrow	NP	VP	1.0
VP	\Rightarrow	Vi		0.4
VP	\Rightarrow	Vt	NP	0.4
VP	\Rightarrow	VP	PP	0.2
NP	\Rightarrow	DT	NN	0.3
NP	\Rightarrow	NP	PP	0.7
PP	\Rightarrow	P	NP	1.0

Vi	\Rightarrow	sleeps	1.0
Vt	\Rightarrow	saw	1.0
NN	\Rightarrow	man	0.7
NN	\Rightarrow	woman	0.2
NN	\Rightarrow	telescope	0.1
DT	\Rightarrow	the	1.0
IN	\Rightarrow	with	0.5
IN	\Rightarrow	in	0.5

Notation Summary

- ▶ $\mathcal{T}(i, j, X)$. Special cases
 - ▶ $\mathcal{T}(1, n, S)$
 - ▶ $\mathcal{T}(i, i, X)$
- ▶ $\pi(i, j, X)$. Special cases
 - ▶ $\pi(1, n, S)$
 - ▶ $\pi(i, i, X)$

Notation Summary

- ▶ $\mathcal{T}(i, j, X)$. Special cases
 - ▶ $\mathcal{T}(1, n, S)$
 - ▶ $\mathcal{T}(i, i, X)$
- ▶ $\pi(i, j, X)$. Special cases
 - ▶ $\pi(1, n, S)$
 - ▶ $\pi(i, i, X)$
- ▶ Parsing:
 - ▶ from $\mathcal{T}(1, n, S)$, find the tree with score $\pi(1, n, S)$
 - ▶ starting points $\mathcal{T}(i, i, X), \forall i \in \{1, \dots, n\}$

An Example

S → NP VP 1.0

PP → P NP 1.0

VP → V NP 0.7

VP → VP PP 0.3

P → *with* 1.0

V → *saw* 1.0

NP → NP PP 0.4

NP → *astronomers* 0.1

NP → *ears* 0.18

NP → *saw* 0.04

NP → *stars* 0.18

NP → *telescopes* 0.1

An Example

$S \rightarrow NP VP$	1.0	$NP \rightarrow NP PP$	0.4
$PP \rightarrow P NP$	1.0	$NP \rightarrow \textit{astronomers}$	0.1
$VP \rightarrow V NP$	0.7	$NP \rightarrow \textit{ears}$	0.18
$VP \rightarrow VP PP$	0.3	$NP \rightarrow \textit{saw}$	0.04
$P \rightarrow \textit{with}$	1.0	$NP \rightarrow \textit{stars}$	0.18
$V \rightarrow \textit{saw}$	1.0	$NP \rightarrow \textit{telescopes}$	0.1

Sentence

astronomers saw stars with ears

Parse Chart

Parsing:

- ▶ from $\mathcal{T}(1, n, S)$, find the tree with score $\pi(1, n, S)$
- ▶ starting points $\mathcal{T}(i, i, X), \forall i \in \{1, \dots, n\}$

	1	2	3	4	5	
astronomers						1
	saw					2
		stars				3
			with			4
				ears		5

Parse Chart

Parsing:

- ▶ from $\mathcal{T}(1, n, S)$, find the tree with score $\pi(1, n, S)$
- ▶ starting points $\mathcal{T}(i, i, X), \forall i \in \{1, \dots, n\}$

	1	2	3	4	5	
astronomers	(1, 1, X)					1
saw		(2, 2, X)				2
stars			(3, 3, X)			3
with				(4, 4, X)		4
ears					(5, 5, X)	5

Parse Chart

Parsing:

- ▶ from $\mathcal{T}(1, n, S)$, find the tree with score $\pi(1, n, S)$
- ▶ starting points $\mathcal{T}(i, i, X), \forall i \in \{1, \dots, n\}$

	1	2	3	4	5	
astronomers	(1, 1, X)				(1, 5, S)	1
saw		(2, 2, X)				2
stars			(3, 3, X)			3
with				(4, 4, X)		4
ears					(5, 5, X)	5

Probabilistic CKY: Base cases

Sentence

astronomers saw stars with ears

$S \rightarrow NP VP$	1.0	$NP \rightarrow NP PP$	0.4
$PP \rightarrow P NP$	1.0	$NP \rightarrow \textit{astronomers}$	0.1
$VP \rightarrow V NP$	0.7	$NP \rightarrow \textit{ears}$	0.18
$VP \rightarrow VP PP$	0.3	$NP \rightarrow \textit{saw}$	0.04
$P \rightarrow \textit{with}$	1.0	$NP \rightarrow \textit{stars}$	0.18
$V \rightarrow \textit{saw}$	1.0	$NP \rightarrow \textit{telescopes}$	0.1

- For $i \in \{1, \dots, n\}$ and for each $X \in \mathcal{N}$

$$\pi(i, i, X) = P(x_i \mid X)$$

Probabilistic CKY: Base cases

Sentence

astronomers saw stars with ears

$S \rightarrow NP VP$	1.0	$NP \rightarrow NP PP$	0.4
$PP \rightarrow P NP$	1.0	$NP \rightarrow \textit{astronomers}$	0.1
$VP \rightarrow V NP$	0.7	$NP \rightarrow \textit{ears}$	0.18
$VP \rightarrow VP PP$	0.3	$NP \rightarrow \textit{saw}$	0.04
$P \rightarrow \textit{with}$	1.0	$NP \rightarrow \textit{stars}$	0.18
$V \rightarrow \textit{saw}$	1.0	$NP \rightarrow \textit{telescopes}$	0.1

- ▶ For $i \in \{1, \dots, n\}$ and for each $X \in \mathcal{N}$

$$\pi(i, i, X) = P(x_i \mid X)$$

- ▶ Example: $x_2 = \textit{saw}$

$$\begin{aligned}\pi(2, 2, V) &= P(V \rightarrow \textit{saw}) = 1.0 \\ \pi(2, 2, NP) &= P(NP \rightarrow \textit{saw}) = 0.04\end{aligned}\tag{5}$$

Probabilistic CKY: Base cases

$S \rightarrow NP VP$	1.0	$NP \rightarrow NP PP$	0.4
$PP \rightarrow P NP$	1.0	$NP \rightarrow \textit{astronomers}$	0.1
$VP \rightarrow V NP$	0.7	$NP \rightarrow \textit{ears}$	0.18
$VP \rightarrow VP PP$	0.3	$NP \rightarrow \textit{saw}$	0.04
$P \rightarrow \textit{with}$	1.0	$NP \rightarrow \textit{stars}$	0.18
$V \rightarrow \textit{saw}$	1.0	$NP \rightarrow \textit{telescopes}$	0.1

	1	2	3	4	5	
astronomers	NP, 0.1					1
saw		V, 1.0 NP, 0.04				2
		stars	NP, 0.18			3
			with	P, 1.0		4
				ears	NP, 0.18	5

Probabilistic CKY: Recursive cases

For each i, j such that $1 \leq i < j \leq n$ and each $X \in \mathcal{N}$

$$\pi(i, j, X) = \max_{Y, Z \in \mathcal{N}, k \in \{i, \dots, j-1\}} P(X \rightarrow Y Z) \cdot \pi(i, k, Y) \cdot \pi(k + 1, j, Z) \quad (6)$$

Probabilistic CKY: Recursive cases

For each i, j such that $1 \leq i < j \leq n$ and each $X \in \mathcal{N}$

$$\pi(i, j, X) = \max_{Y, Z \in \mathcal{N}, k \in \{i, \dots, j-1\}} P(X \rightarrow Y Z) \cdot \pi(i, k, Y) \cdot \pi(k + 1, j, Z) \quad (6)$$

Example with $i = 2$ and $j = 3$

$$\pi(2, 3, X) = \max_{Y, Z \in \mathcal{N}, k \in \{2\}} P(X \rightarrow Y Z) \cdot \pi(2, k, Y) \cdot \pi(k + 1, 3, Z) \quad (7)$$

Probabilistic CKY: Recursive cases

Example with $i = 2$ and $j = 3$

$$\pi(2, 3, X) = \max_{Y, Z \in \mathcal{N}, k \in \{2\}} P(X \rightarrow Y Z) \cdot \pi(2, k, Y) \cdot \pi(k + 1, 3, Z) \quad (8)$$

► $\pi(2, 3, \text{NP}) = P(\text{NP} \rightarrow Y Z) \cdot \pi(2, 2, Y) \cdot \pi(3, 3, Z), Y, Z \in \mathcal{N}$

$S \rightarrow \text{NP VP}$	1.0	$\text{NP} \rightarrow \text{NP PP}$	0.4
$\text{PP} \rightarrow \text{P NP}$	1.0	$\text{NP} \rightarrow \text{astronomers}$	0.1
$\text{VP} \rightarrow \text{V NP}$	0.7	$\text{NP} \rightarrow \text{ears}$	0.18
$\text{VP} \rightarrow \text{VP PP}$	0.3	$\text{NP} \rightarrow \text{saw}$	0.04
$\text{P} \rightarrow \text{with}$	1.0	$\text{NP} \rightarrow \text{stars}$	0.18
$\text{V} \rightarrow \text{saw}$	1.0	$\text{NP} \rightarrow \text{telescopes}$	0.1

	1	2	3	4	5	
astronomers	NP, 0.1					1
saw		V, 1.0 NP, 0.04				2
		stars	NP, 0.18			3
			with	P, 1.0		4
				ears	NP, 0.18	5

Probabilistic CKY: Recursive cases

Example with $i = 2$ and $j = 3$

$$\pi(2, 3, X) = \max_{Y, Z \in \mathcal{N}, k \in \{2\}} P(X \rightarrow Y Z) \cdot \pi(2, k, Y) \cdot \pi(k + 1, 3, Z) \quad (8)$$

- ▶ $\pi(2, 3, \text{NP}) = P(\text{NP} \rightarrow Y Z) \cdot \pi(2, 2, Y) \cdot \pi(3, 3, Z), Y, Z \in \mathcal{N}$
- ▶ $\pi(2, 3, \text{VP}) = P(\text{VP} \rightarrow Y Z) \cdot \pi(2, 2, Y) \cdot \pi(3, 3, Z), Y, Z \in \mathcal{N}$

$S \rightarrow \text{NP VP}$	1.0	$\text{NP} \rightarrow \text{NP PP}$	0.4
$\text{PP} \rightarrow \text{P NP}$	1.0	$\text{NP} \rightarrow \text{astronomers}$	0.1
$\text{VP} \rightarrow \text{V NP}$	0.7	$\text{NP} \rightarrow \text{ears}$	0.18
$\text{VP} \rightarrow \text{VP PP}$	0.3	$\text{NP} \rightarrow \text{saw}$	0.04
$\text{P} \rightarrow \text{with}$	1.0	$\text{NP} \rightarrow \text{stars}$	0.18
$\text{V} \rightarrow \text{saw}$	1.0	$\text{NP} \rightarrow \text{telescopes}$	0.1

	1	2	3	4	5	
astronomers	NP, 0.1					1
saw		V, 1.0 NP, 0.04				2
stars			NP, 0.18			3
with				P, 1.0		4
ears					NP, 0.18	5

Recursive Cases: Example (I)

$$\pi(2, 3, X) = \max_{Y, Z \in \mathcal{N}, k \in \{2\}} P(X \rightarrow Y Z) \cdot \pi(2, k, Y) \cdot \pi(k + 1, 3, Z) \quad (9)$$

S → NP VP	1.0	NP → NP PP	0.4
PP → P NP	1.0	NP → <i>astronomers</i>	0.1
VP → V NP	0.7	NP → <i>ears</i>	0.18
VP → VP PP	0.3	NP → <i>saw</i>	0.04
P → <i>with</i>	1.0	NP → <i>stars</i>	0.18
V → <i>saw</i>	1.0	NP → <i>telescopes</i>	0.1

	1	2	3	4	5	
astronomers	NP, 0.1					1
saw		V, 1.0 NP, 0.04	VP, 0.126			2
		stars	NP, 0.18			3
			with	P, 1.0		4
				ears	NP, 0.18	5

Recursive Cases: Example (II)

$$\pi(1, 2, X) = \max_{L, R \in \mathcal{N}, k \in \{1\}} P(X \rightarrow Y Z) \cdot \pi(1, 1, Y) \cdot \pi(2, 2, Z) \quad (10)$$

$S \rightarrow NP VP$	1.0	$NP \rightarrow NP PP$	0.4
$PP \rightarrow P NP$	1.0	$NP \rightarrow \textit{astronomers}$	0.1
$VP \rightarrow V NP$	0.7	$NP \rightarrow \textit{ears}$	0.18
$VP \rightarrow VP PP$	0.3	$NP \rightarrow \textit{saw}$	0.04
$P \rightarrow \textit{with}$	1.0	$NP \rightarrow \textit{stars}$	0.18
$V \rightarrow \textit{saw}$	1.0	$NP \rightarrow \textit{telescopes}$	0.1

	1	2	3	4	5	
astronomers	NP, 0.1	\emptyset				1
saw		$V, 1.0$ $NP, 0.04$	VP, 0.126			2
		stars	NP, 0.18			3
			with	P, 1.0		4
				ears	NP, 0.18	5

Recursive Cases

$$\pi(i, j, X) = \max_{Y, Z \in \mathcal{N}, k \in \{i, \dots, j-1\}} P(X \rightarrow Y Z) \cdot \pi(i, k, Y) \cdot \pi(k+1, j, Z)$$

S → NP VP	1.0	NP → NP PP	0.4
PP → P NP	1.0	NP → <i>astronomers</i>	0.1
VP → V NP	0.7	NP → <i>ears</i>	0.18
VP → VP PP	0.3	NP → <i>saw</i>	0.04
P → <i>with</i>	1.0	NP → <i>stars</i>	0.18
V → <i>saw</i>	1.0	NP → <i>telescopes</i>	0.1

	1	2	3	4	5	
astronomers	NP, 0.1	∅				1
saw		V, 1.0 NP, 0.04	VP, 0.126			2
		stars	NP, 0.18	∅		3
			with	P, 1.0	PP, 0.18	4
				ears	NP, 0.18	5

Probabilistic CKY: Recursive cases

$$\pi(i, j, X) = \max_{Y, Z \in \mathcal{N}, k \in \{i, \dots, j-1\}} P(X \rightarrow Y Z) \cdot \pi(i, k, Y) \cdot \pi(k+1, j, Z)$$

S → NP VP	1.0	NP → NP PP	0.4
PP → P NP	1.0	NP → <i>astronomers</i>	0.1
VP → V NP	0.7	NP → <i>ears</i>	0.18
VP → VP PP	0.3	NP → <i>saw</i>	0.04
P → <i>with</i>	1.0	NP → <i>stars</i>	0.18
V → <i>saw</i>	1.0	NP → <i>telescopes</i>	0.1

	1	2	3	4	5	
astronomers	NP, 0.1	∅	S, 0.0126			1
saw		V, 1.0 NP, 0.04	VP, 0.126	∅		2
		stars	NP, 0.18	∅	NP, 0.01296	3
			with	P, 1.0	PP, 0.18	4
				ears	NP, 0.18	5

Probabilistic CKY: Recursive cases

$$\pi(i, j, X) = \max_{Y, Z \in \mathcal{N}, k \in \{i, \dots, j-1\}} P(X \rightarrow Y Z) \cdot \pi(i, k, Y) \cdot \pi(k+1, j, Z)$$

S → NP VP	1.0	NP → NP PP	0.4
PP → P NP	1.0	NP → <i>astronomers</i>	0.1
VP → V NP	0.7	NP → <i>ears</i>	0.18
VP → VP PP	0.3	NP → <i>saw</i>	0.04
P → <i>with</i>	1.0	NP → <i>stars</i>	0.18
V → <i>saw</i>	1.0	NP → <i>telescopes</i>	0.1

	1	2	3	4	5	
astronomers	NP, 0.1	∅	S, 0.0126	∅		1
saw		V, 1.0 NP, 0.04	VP, 0.126	∅	VP, 0.015876	2
		stars	NP, 0.18	∅	NP, 0.01296	3
			with	P, 1.0	PP, 0.18	4
				ears	NP, 0.18	5

Probabilistic CKY: Recursive cases

$$\pi(i, j, X) = \max_{Y, Z \in \mathcal{N}, k \in \{i, \dots, j-1\}} P(X \rightarrow Y Z) \cdot \pi(i, k, Y) \cdot \pi(k+1, j, Z)$$

S → NP VP	1.0	NP → NP PP	0.4
PP → P NP	1.0	NP → <i>astronomers</i>	0.1
VP → V NP	0.7	NP → <i>ears</i>	0.18
VP → VP PP	0.3	NP → <i>saw</i>	0.04
P → <i>with</i>	1.0	NP → <i>stars</i>	0.18
V → <i>saw</i>	1.0	NP → <i>telescopes</i>	0.1

	1	2	3	4	5	
astronomers	NP, 0.1	∅	S, 0.0126	∅	S, 0.0015876	1
saw		V, 1.0 NP, 0.04	VP, 0.126	∅	VP, 0.015876	2
		stars	NP, 0.18	∅	NP, 0.01296	3
			with	P, 1.0	PP, 0.18	4
				ears	NP, 0.18	5

Input: a sentence $s = x_1 \dots x_n$, a PCFG $G = (N, \Sigma, S, R, q)$.

Initialization:

For all $i \in \{1 \dots n\}$, for all $X \in N$,

$$\pi(i, i, X) = \begin{cases} q(X \rightarrow x_i) & \text{if } X \rightarrow x_i \in R \\ 0 & \text{otherwise} \end{cases}$$

Algorithm:

- For $l = 1 \dots (n - 1)$
 - For $i = 1 \dots (n - l)$
 - * Set $j = i + l$
 - * For all $X \in N$, calculate

$$\pi(i, j, X) = \max_{\substack{X \rightarrow YZ \in R, \\ s \in \{i \dots (j-1)\}}} (q(X \rightarrow YZ) \times \pi(i, s, Y) \times \pi(s + 1, j, Z))$$

and

$$bp(i, j, X) = \arg \max_{\substack{X \rightarrow YZ \in R, \\ s \in \{i \dots (j-1)\}}} (q(X \rightarrow YZ) \times \pi(i, s, Y) \times \pi(s + 1, j, Z))$$

Output: Return $\pi(1, n, S) = \max_{t \in \mathcal{T}(s)} p(t)$, and backpointers bp which allow recovery of $\arg \max_{t \in \mathcal{T}(s)} p(t)$.

- ▶ Space and runtime requirements

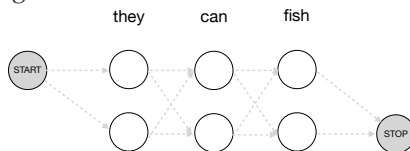
- ▶ Space: $\mathcal{O}(|\mathcal{N}|n^2)$

- ▶ Time: $\mathcal{O}(|\mathcal{N}|n^3)$

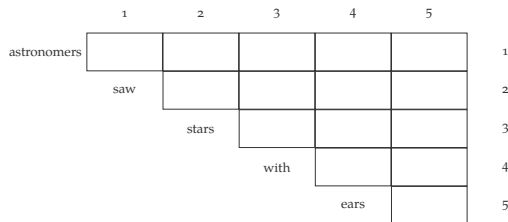
astronomers					
	saw				
		stars			
			with		
				ears	

Probabilistic CKY vs. Viterbi Decoding

► Viterbi decoding



► Probabilistic CKY



Keywords: conditional independence, forward enumerating, backward tracing, dynamic programming

1. Probabilistic CFGs
2. Probabilistic CKY Algorithm

Reference



Collins, M. (2017).

Natural language processing: Lecture notes.



Smith, N. A. (2017).

Natural language processing: Lecture notes.