

PCFGs

Scott Farrar
CLMA, University of Washington
farrar@u.washington.edu

January 20, 2010

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Today's lecture

- 1 Uncertainty
- 2 Probabilistic Context-Free Grammars
 - Uses of PCFGs: Disambiguation
 - Benefits of PCFGs
 - Issues

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

*To get out of recession, we must **obamafy** our economy.*

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

*To get out of recession, we must **obamafy** our economy.*

When the input is unknown, then you're dealing with a **random variable**.

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

*To get out of recession, we must **obamafy** our economy.*

When the input is unknown, then you're dealing with a **random variable**.

- the weather

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

*To get out of recession, we must **obamafy** our economy.*

When the input is unknown, then you're dealing with a **random variable**.

- the weather
- the stock market

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

*To get out of recession, we must **obamafy** our economy.*

When the input is unknown, then you're dealing with a **random variable**.

- the weather
- the stock market
- what someone will say next to a speech recognition system

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Uncertain but regular

Strategy for leveraging probability theory for NLP: exploit regularities in known data to predict the structure of unseen data.

Uncertain but regular

Strategy for leveraging probability theory for NLP: exploit regularities in known data to predict the structure of unseen data.

Example

To get out of recession, we must ?????? our economy.

- We can exploit syntax.

Uncertain but regular

Strategy for leveraging probability theory for NLP: exploit regularities in known data to predict the structure of unseen data.

Example

To get out of recession, we must ?????? our economy.

- We can exploit syntax.
- We can exploit morphology.

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Uncertain but regular

Strategy for leveraging probability theory for NLP: exploit regularities in known data to predict the structure of unseen data.

Example

To get out of recession, we must ??????fy our economy.

- We can exploit syntax.
- We can exploit morphology.

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Today's lecture

1 Uncertainty

2 Probabilistic Context-Free Grammars

- Uses of PCFGs: Disambiguation
- Benefits of PCFGs
- Issues

What is a PCFG?

Definition

Also known as a *weighted* grammar, a probabilistic context-free grammar (PCFG) is one that assigns a probability to each production rule.

What is a PCFG?

Definition

Also known as a *weighted* grammar, a probabilistic context-free grammar (PCFG) is one that assigns a probability to each production rule.

It can be used to assign a probability to every string in the language (language model) and to every structure in the language.

What is a PCFG?

Definition

Also known as a *weighted* grammar, a probabilistic context-free grammar (PCFG) is one that assigns a probability to each production rule.

It can be used to assign a probability to every string in the language (language model) and to every structure in the language.

S = *To get out of recession, we must obamafy the economy.*
 $P(\mathbf{S}) = 0.000000004538$

Uncertainty

Probabilistic
Context-Free
GrammarsUses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Formal definition of a PCFG

Definition

A CFG grammar is formally defined as $G = \langle N, \Sigma, P, S \rangle$
where:

- N is a set of non-terminal symbols, typically S, A, B, \dots
- S is the starting or goal symbol from N , i.e., $S \in N$
- Σ is a set of terminal symbols, typically x, y, z, \dots
disjoint from N
- P is a set of production rules with attached probabilities.

Formal definition of a PCFG (cont)

Definition

The productions P are of the form: $A \rightarrow \beta [p]$, where:

- A is a non-terminal $A \in N$
- β is a string of symbols from $(\Sigma \cup N)$
- $[p]$ is the probability (from 0 to 1) that A will have β as its constituent.

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2 : $P(e_1|e_2)$, e.g., $P(\text{raining}|\text{wetroads})$.

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2 : $P(e_1|e_2)$, e.g., $P(\text{raining}|\text{wetroads})$.
- In general then, we can express the conditional probability for any context-free rule in the following way: $P(A \rightarrow \beta)$ or $P(A \rightarrow \beta|A)$ or simply $P(\beta|A)$.

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2 : $P(e_1|e_2)$, e.g., $P(\text{raining}|\text{wetroads})$.
- In general then, we can express the conditional probability for any context-free rule in the following way: $P(A \rightarrow \beta)$ or $P(A \rightarrow \beta|A)$ or simply $P(\beta|A)$.
- In general then: conditional probability for any context-free rule: $P(\text{RHS}|\text{LHS})$

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2 : $P(e_1|e_2)$, e.g., $P(\text{raining}|\text{wetroads})$.
- In general then, we can express the conditional probability for any context-free rule in the following way: $P(A \rightarrow \beta)$ or $P(A \rightarrow \beta|A)$ or simply $P(\beta|A)$.
- In general then: conditional probability for any context-free rule: $P(\text{RHS}|\text{LHS})$

PCFGs are defined according to a generative perspective:
top-down

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2 : $P(e_1|e_2)$, e.g., $P(\text{raining}|\text{wetroads})$.
- In general then, we can express the conditional probability for any context-free rule in the following way: $P(A \rightarrow \beta)$ or $P(A \rightarrow \beta|A)$ or simply $P(\beta|A)$.
- In general then: conditional probability for any context-free rule: $P(\text{RHS}|\text{LHS})$

PCFGs are defined according to a generative perspective:
top-down

It turns out that the non-generative approach, i.e.,
 $P(\text{LHS}|\text{RHS})$, gives a poor parsing model.

Calculating probabilities

Treebank as grammar

A rule probability is calculated by summing the occurrences of rules in a labeled, bracketed treebank. That is, the treebank acts a kind of implicit grammar, and the explicit grammar can be “induced” from labeled and bracketed trees.

$$P(A \rightarrow \beta | A) = \frac{\text{Count}(A \rightarrow \beta)}{\sum_{\gamma} \text{Count}(A \rightarrow \gamma)} = \frac{\text{Count}(A \rightarrow \beta)}{\text{Count}(A)}$$

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times
$S \rightarrow NP VP$	occurs 97 times
$S \rightarrow AUX NP VP$	occurs 43
$S \rightarrow VP$	occurs 58
$S \rightarrow ADJP$	occurs 14 times

Uncertainty

Probabilistic
 Context-Free
 Grammars

Uses of PCFGs:
 Disambiguation
 Benefits of PCFGs
 Issues

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP VP$	occurs 97 times	$P=0.4575$
$S \rightarrow AUX NP VP$	occurs 43	
$S \rightarrow VP$	occurs 58	
$S \rightarrow ADJP$	occurs 14 times	

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP VP$	occurs 97 times	$P=0.4575$
$S \rightarrow AUX NP VP$	occurs 43	$P=0.2028$
$S \rightarrow VP$	occurs 58	
$S \rightarrow ADJP$	occurs 14 times	

Uncertainty

Probabilistic
 Context-Free
 Grammars

Uses of PCFGs:
 Disambiguation
 Benefits of PCFGs
 Issues

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP VP$	occurs 97 times	$P=0.4575$
$S \rightarrow AUX NP VP$	occurs 43	$P=0.2028$
$S \rightarrow VP$	occurs 58	$P=0.2736$
$S \rightarrow ADJP$	occurs 14 times	

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP VP$	occurs 97 times	$P=0.4575$
$S \rightarrow AUX NP VP$	occurs 43	$P=0.2028$
$S \rightarrow VP$	occurs 58	$P=0.2736$
$S \rightarrow ADJP$	occurs 14 times	$P=0.0660$

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP VP$	occurs 97 times	$P=0.4575$
$S \rightarrow AUX NP VP$	occurs 43	$P=0.2028$
$S \rightarrow VP$	occurs 58	$P=0.2736$
$S \rightarrow ADJP$	occurs 14 times	$P=0.0660$
		$0.9999 \approx 1$

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Using PCFGs for disambiguation

Probability of a particular parse tree (its derivation) is defined as:

$$P(T) = \prod_{i=1}^n P(RHS_i | LHS_i)$$

Using PCFGs for disambiguation

Probability of a particular parse tree (its derivation) is defined as:

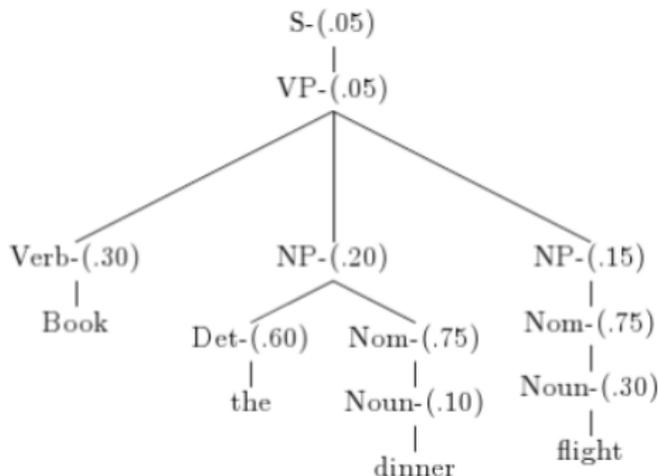
$$P(T) = \prod_{i=1}^n P(RHS_i | LHS_i)$$

Using PCFGs for disambiguation

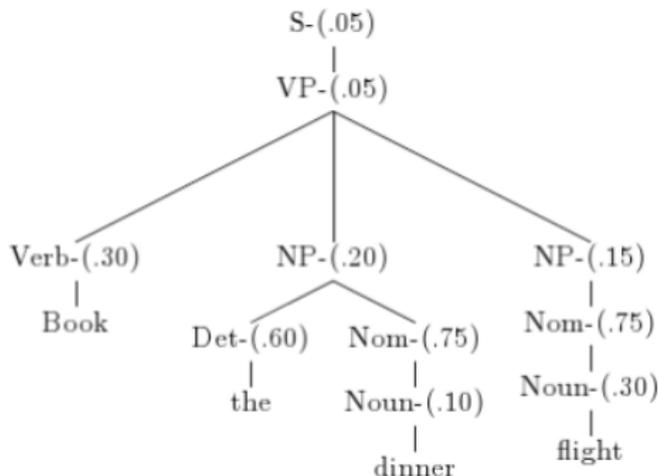
Probability of a particular parse tree (its derivation) is defined as:

$$P(T) = \prod_{i=1}^n P(RHS_i | LHS_i)$$

That is, we multiply together all rule expansion probabilities to obtain the probability of the tree.



$$.05 * .05 * .30 * .20 * .60 * .75 * .10 * .15 * .75 * .30 = 2.28 \times 10^{-7}$$



$$.05 * .05 * .30 * .20 * .60 * .75 * .10 * .15 * .75 * .30 = 2.28 \times 10^{-7}$$

Parse tree 1 was greater: $1.62 \times 10^{-6} > 2.28 \times 10^{-7}$

Disambiguation strategy

A disambiguation strategy is to choose the parse with the highest probability:

$$\hat{T}(S) = \arg \max_{\text{yield}(T)}$$

- Choose the most likely parse \hat{T} given the $\text{yield}(T)$, or the string sequence S
- What exactly is being maximized?
- The probability of the parse tree T given input sentence S .

Uncertainty

Probabilistic
Context-Free
GrammarsUses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Benefits of using PCFGs

Scott Farrar
CLMA, University
of Washington far-
rar@u.washington.edu

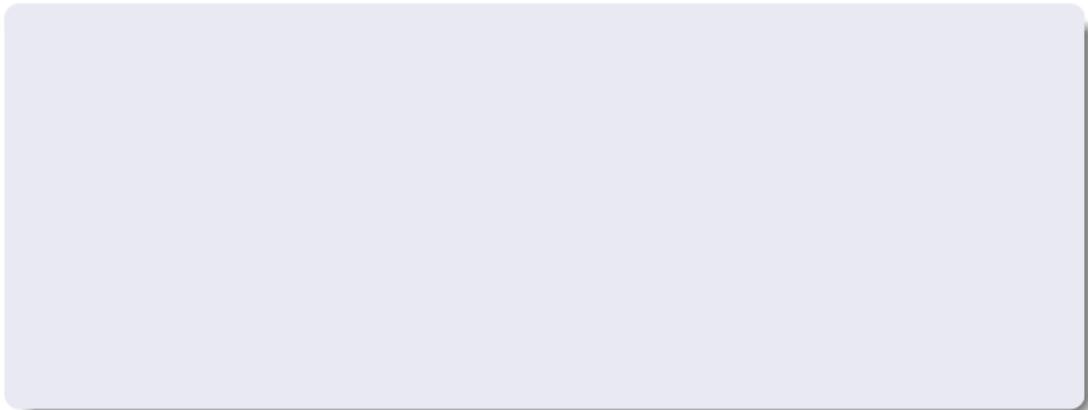
Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation

Benefits of PCFGs

Issues



Benefits of using PCFGs

- As the number of possible trees for a given input grows, a PCFG gives some idea of the plausibility of a particular parse.

Benefits of using PCFGs

- As the number of possible trees for a given input grows, a PCFG gives some idea of the plausibility of a particular parse.
- PCFGs help smooth out disfluencies and other errors in the training set.

Benefits of using PCFGs

- As the number of possible trees for a given input grows, a PCFG gives some idea of the plausibility of a particular parse.
- PCFGs help smooth out disfluencies and other errors in the training set.
- PCFGs are good for inducing grammars from positive data (from parsed treebanks).

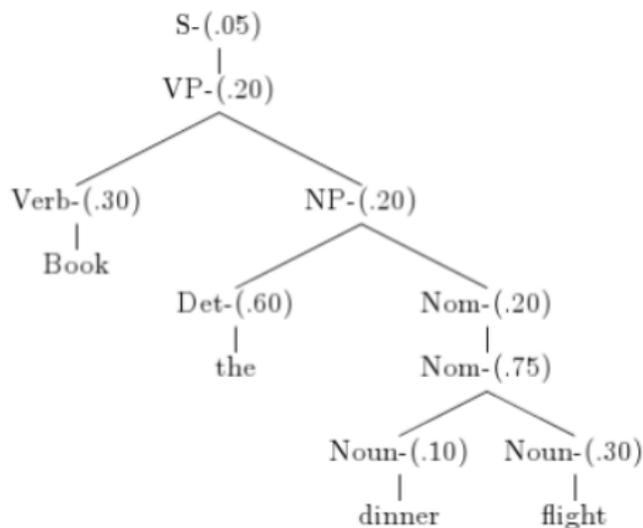
Place invariance

probability of a subtree does not depend on the position of its terminals

Place invariance

probability of a subtree does not depend on the position of its terminals

Recall example tree:



$$P(NP) = .20 \times .60 \times .20 \times .75 \times .10 \times .30 = 0.00054$$

Some assumptions with PCFGs

- **place invariance:** probability of a subtree does not depend on the position of its terminals

Some assumptions with PCFGs

- **place invariance:** probability of a subtree does not depend on the position of its terminals
 $P(\textit{the dog bit the boy}) = P(\textit{the boy bit the dog})$

Some assumptions with PCFGs

- **place invariance:** probability of a subtree does not depend on the position of its terminals
 $P(\textit{the dog bit the boy}) = P(\textit{the boy bit the dog})$
- **context-free:** probability of a subtree does not depend on words not dominated by the subtree

Some assumptions with PCFGs

- **place invariance:** probability of a subtree does not depend on the position of its terminals
 $P(\textit{the dog bit the boy}) = P(\textit{the boy bit the dog})$
- **context-free:** probability of a subtree does not depend on words not dominated by the subtree
(the bird) swims ..., *(the bird) flies ...*

Context sensitivity of certain rules

Are subjects usually pronouns or full NPs? what about objects?

Context sensitivity of certain rules

Are subjects usually pronouns or full NPs? what about objects?

	Pronoun	Non-pronoun
Subject	91%	9%
Object	34%	66%

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Context sensitivity of certain rules

Are subjects usually pronouns or full NPs? what about objects?

	Pronoun	Non-pronoun
Subject	91%	9%
Object	34%	66%

Example

He saw the accident involving the red car.

Uncertainty

Probabilistic
Context-Free
Grammars

Uses of PCFGs:
Disambiguation
Benefits of PCFGs
Issues

Context sensitivity of certain rules

Are subjects usually pronouns or full NPs? what about objects?

	Pronoun	Non-pronoun
Subject	91%	9%
Object	34%	66%

Example

He saw the accident involving the red car.

should be more likely than:

Example

The man in the street near the bank saw it