Simple Probabilistic Modeling and PP Attachment

Ambiguity

- I saw the dog with the blue hat
- He talked to the girl in a harsh voice
- Graucho shot an elephant in his pajamas
- John gave Mary a sack of money
- He thought about filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill with a telescope

Ambiguity

- I saw the dog with the blue hat
- He talked to the girl in a harsh voice
- One morning, Graucho shot an elephant in his pajamas
- John found a sack of money
- He thought about filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill with a telescope

These are all the same (how?)

Ambiguity

- I saw the dog with the blue hat
- He talked to the girl in a harsh voice
- Graucho shot an elephant in his pajamas
- John found a sack of money
- He thought about filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill with a telescope

These are all the same (how?)

Ambiguity ??

- I saw the dog with the blue hat
- He talked to the girl in a harsh voice
- Graucho shot an elephant in his pajamas
- John found a sack of money
- He **thought about** filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill **with a telescope**

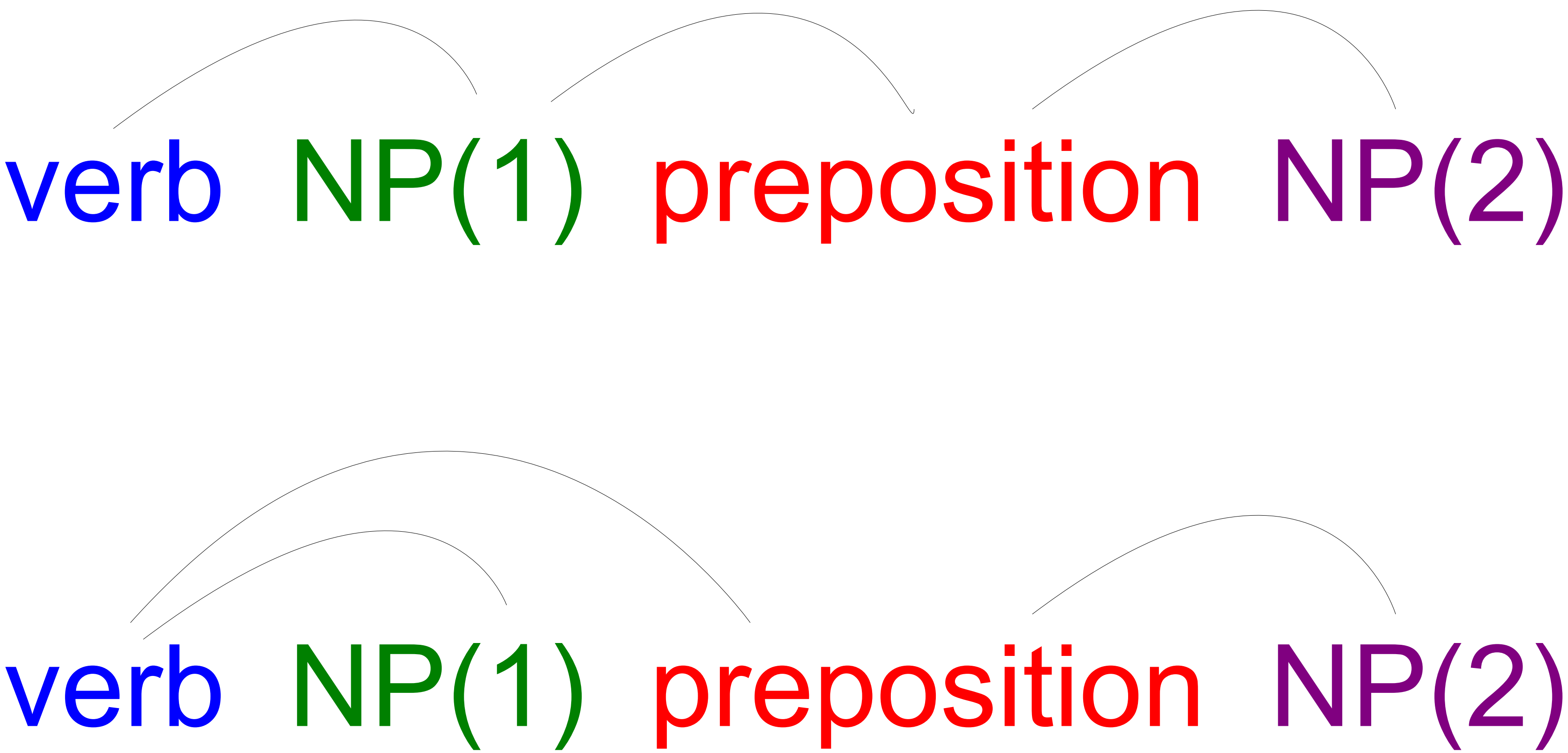
These are all the same (how?)

Ambiguity

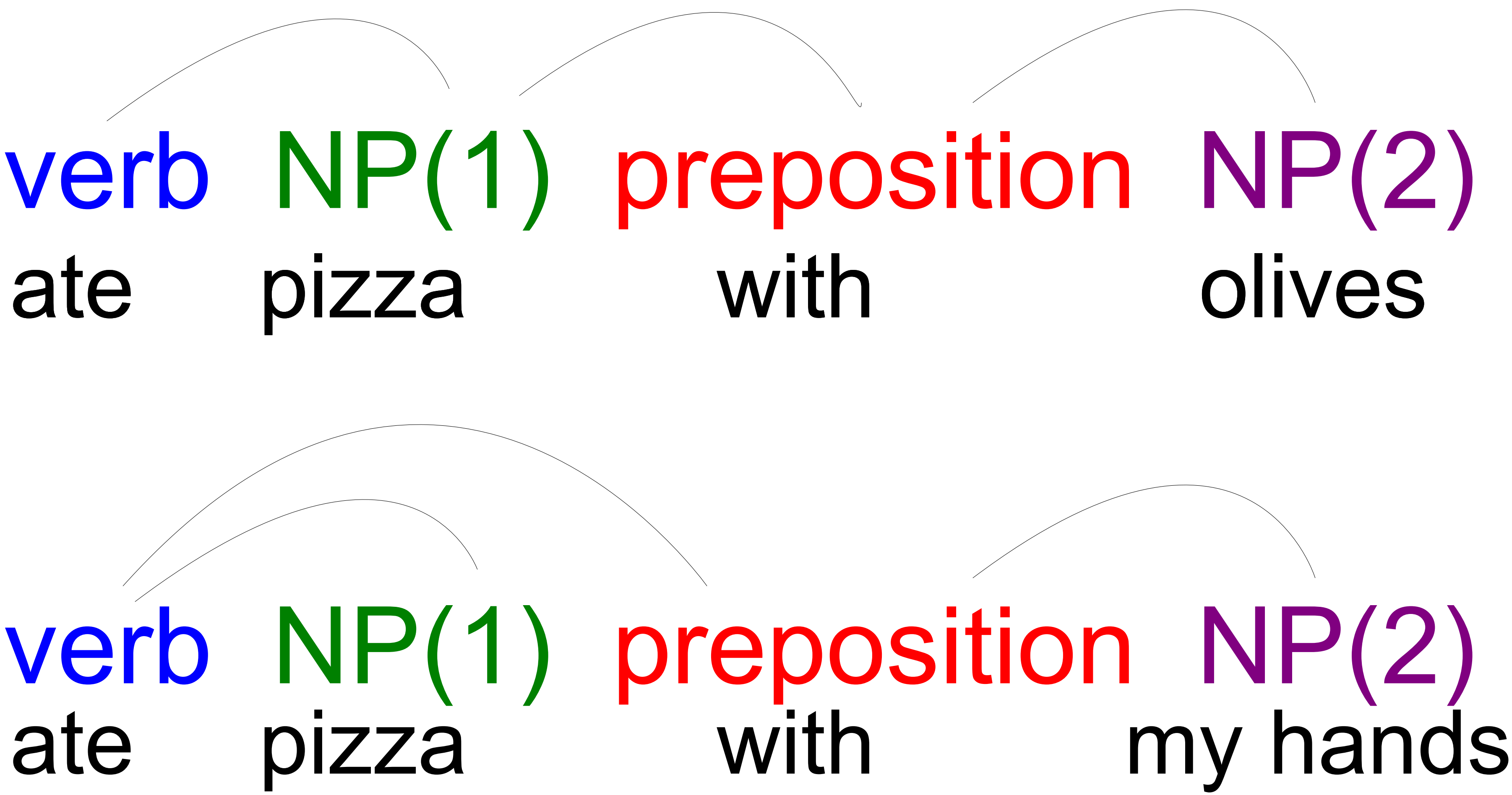
- I saw the dog with the blue hat
- He talked to the girl in a harsh voice
- Graucho shot an elephant in his pajamas
- John found a sack of money
- He thought about filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill with a telescope

verb NP(1) preposition NP(2)

Ambiguity



Ambiguity


verb NP(1) preposition NP(2)
ate pizza with olives
verb NP(1) preposition NP(2)
ate pizza with my hands

The N-V PP attachment problem

- You get a 4-tuple: (verb, NP1, prep, NP2)
  - talked the girl in a harsh voice
  - shot an elephant in his pajamas
  - found a sack of money
  - filling the garden with flowers
- Need to decide: V or N
  - V means a V-PREP relation (ate with my hands)
  - N means a N-PREP relation (pizza with olives)
- A **binary classification** task

The N-V PP attachment problem

- You get a 4-tuple: (verb, NP1, prep, NP2)
  - talked the girl in a harsh voice
  - shot an elephant in his pajamas
  - found a sack of money
  - filling the garden with flowers
- Need to decide: V or N
  - V means a V-PREP relation (ate with my hands)
  - N means a N-PREP relation (pizza with olives)
- A **binary classification** task

Where do the tuples come from???

One morning I shot an elephant in my pajamas.
How he got into my pajamas I'll never know.

- Graucho Marx

Sometimes, must use discourse...

Ambiguity

- I saw the dog with the blue hat
- He talked to the girl in a harsh voice
- Graucho shot an elephant in his pajamas
- John found a sack of money
- He thought about filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill with a telescope

verb NP(1) preposition NP(2)

Ambiguity

- I saw the dog with the blue hat
- He talked to the girl in a harsh voice
- Graucho shot an elephant in his pajamas
- John found a sack of money
- He thought about filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill with a telescope

verb NP(1) preposition NP(2)

Ambiguity

- I saw the dog with the blue hat
- He talked to the girl in a harsh voice
- Graucho shot an elephant in his pajamas
- John found a sack of money
- He thought about filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill with a telescope

verb noun(1) preposition noun(2)

Modeling choice:
consider only the head (“main”) words

Is this a reasonable thing to do?
why?
why not?
(what do we gain? what do we lose?)

- Graucho shot an elephant in his pajamas
- John found a sack of money
- He thought about filling the garden with flowers
- Collect the young children after school
- I saw a boy on the hill with a telescope

verb noun(1) preposition noun(2)

The N-V PP attachment problem

- You get a 4-tuple: (verb, noun1, prep, noun2)
  - talked girl in voice
  - shot elephant in pajamas
  - found sack of money
  - filling garden with flowers
- Need to decide: V or N
  - V means a V-PREP relation (ate with hands)
  - N means a N-PREP relation (pizza with olives)
- A **binary classification** task

How do we solve it?

- Assume supervised classification:
  - You get 4000 (or 40,000, or 400,000) tuples with their correct answer.
    - talked girl in voice → V
    - shot elephant in pajamas → V
    - found sack of money → N
    - filling garden with flowers → V
    - ...
  - Someone hands new a new tuple. Need to decide based on previous observation.

Step 1 (always) → Look at the data

Step 1 (always) → Look at the data
Step 2 (always) → Define accuracy measure

Step 1 (always) → Look at the data

Step 2 (always) → Define accuracy measure

acc = correct / (correct + incorrect)

How do we solve it?

- Conditional probability:

```
if P(V | verb, noun1, prep, noun2) > 0.5
        say V
else:
        say N
```

for example, P(V | saw, boy, with, hat)

Maximum Likelihood Estimation

$$P(\text{ V } | \text{ verb, noun1, prep, noun2}) = \frac{\text{count(V,verb, noun1, prep, noun2)}}{\text{count(*,verb, noun1, prep, noun2)}}$$

count(...) is number of times we saw the event in the training data

- This is called MLE estimation. (maximum likelihood)

Maximum Likelihood Estimation

$$P(\,V \mid verb, noun1, prep, noun2) = \frac{count(V, verb, noun1, prep, noun2)}{count(*, verb, noun1, prep, noun2)}$$

count(...) is number of times we saw the event in the training data

Is this reasonable? Why?

Maximum Likelihood Estimation

$$P(\ V \mid \text{verb, noun1, prep, noun2}) = \frac{\text{count}(V,\text{verb, noun1, prep, noun2})}{\text{count}(*,\text{verb, noun1, prep, noun2})}$$

count(...) is number of times we saw the event in the training data

Problem: data sparsity and overfitting

Another option (majority baseline)

$P(\ V \mid \text{verb, noun1, prep, noun2}) \approx P(V)$

Is this reasonable?

What score would you expect?

Another option

$P(V \mid \text{verb}, \text{noun1}, \text{prep}, \text{noun2}) \approx P(V|\text{noun1})$

Is this reasonable?

What score would you expect?

Another option

$$P(\,V \mid verb, noun1, prep, noun2) \approx P(V|prep)$$

Is this reasonable?

What score what score would you expect?

Another option

$$P(\ V \mid \text{verb, noun1, prep, noun2}) \approx P(V \mid \text{prep})$$

Is this reasonable?

What score what score would you expect?

This one is actually pretty good! (why?)

Another option

$$P(\ V \mid \text{verb, noun1, prep, noun2}) \approx P(V|\text{prep})$$

Is this reasonable?

What score what score would you expect?

This one is actually pretty good! (why?)

Can we do better?

P(V| verb, prep) ?

P(V| noun1, prep) ?

P(V| noun1, noun2) ?

P(V| verb, noun1, noun2) ?

P(V| verb, noun1, prep) ?

How do we combine the different probabilities?

- Remember, for a function to be a probability function, we must have:
  - always positive
  - sum to one

- (do we care if our scoring function is a probability function? why?)

How do we combine the different probabilities?

- One way of combining probabilities to obtain a probability is **linear interpolation**

$$P_{interpolate} = \lambda_1 P_1 + \lambda_2 P_2 + \lambda_3 P_3 ... + \lambda_k P_k$$

$$\lambda_1 + \lambda_2 + \lambda_3 + ... + \lambda_k = 1$$

Collins and Brooks' estimation

- Interpolate

  $P(V|v,n1,p)$, $P(V|v,p,n2)$, $P(V|n1,p,n2)$ into $P_{triplet}$

- Interpolate

  $P(V|v,p)$, $P(V|n1,p)$, $P(V|p,n2)$ into P_{pair}

Collins and Brooks' estimation

- Interpolate

  $P(V|v,n1,p)$, $P(V|v,p,n2)$, $P(V|n1,p,n2)$ into $P_{triplet}$

- Interpolate

  $P(V|v,p)$, $P(V|n1,p)$, $P(V|p,n2)$ into P_{pair}

Notice we always include **p** (the preposition).

We do not have $P(V|n1,n2)$ for example.

Why?

Collins and Brooks' estimation

- Interpolate

  $P(V|v,n1,p)$, $P(V|v,p,n2)$, $P(V|n1,p,n2)$ into $P_{triplet}$

  $P(V|v,n1,p) = \#(V, v, n1, p, *) / \#(*, v, n1, p, *)$

- Interpolate

  $P(V|v,p)$, $P(V|n1,p)$, $P(V|p,n2)$ into P_{pair}

Collins and Brooks' estimation

- Interpolate

  $P(V|v,n1,p)$, $P(V|v,p,n2)$, $P(V|n1,p,n2)$ into $P_{triplet}$

  $P(V|v,n1,p) = \#(V, v, n1, p, *) / \#(*, v, n1, p, *)$

- Interpolate

  $P(V|v,p)$, $P(V|n1,p)$, $P(V|p,n2)$ into P_{pair}

  $P(V|v, p) = \#(V, v, *, p, *) / \#(*, v, *, p, *)$

How do we combine the different probabilities?

- One way of combining probabilities to obtain a probability is **linear interpolation**

$$P_{interpolate} = \lambda_1 P_1 + \lambda_2 P_2 + \lambda_3 P_3 ... + \lambda_k P_k$$

$$\lambda_1 + \lambda_2 + \lambda_3 + ... + \lambda_k = 1$$

Collins and Brooks' interpolation

$$\lambda_{v,n1,p} = \frac{count(v,n1,p)}{count(v,n1,p) + count(v,p,n2) + count(n1,p,n2)}$$

$$\lambda_{v,p,n2} = \frac{count(v,p,n2)}{count(v,n1,p) + count(v,p,n2) + count(n1,p,n2)}$$

$$\lambda_{n1,p,n2} = \frac{count(n1,p,n2)}{count(v,n1,p) + count(v,p,n2) + count(n1,p,n2)}$$

Collins and Brooks' interpolation

$$\lambda_{v,n1,p}=\frac{count(v,n1,p)}{count(v,n1,p)+count(v,p,n2)+count(n1,p,n2)}$$

$$\lambda_{v,p,n2}=\frac{count(v,p,n2)}{count(v,n1,p)+count(v,p,n2)+count(n1,p,n2)}$$

$$\lambda_{n1,p,n2}=\frac{count(n1,p,n2)}{count(v,n1,p)+count(v,p,n2)+count(n1,p,n2)}$$

Give more weight to events that occurred
more times in the training data.

Collins and Brooks' estimation

$$P_3(V|v,n1,p,n2) = \frac{count(V,v,n1,p) + count(V,v,p,n2) + count(V,n1,p,n2)}{count(*,v,n1,p) + count(*,v,p,n2) + count(*,n1,p,n2)}$$

Collins and Brooks' estimation

$$P_3(V|v,n1,p,n2) = \frac{count(V,v,n1,p) + count(V,v,p,n2) + count(V,n1,p,n2)}{count(*,v,n1,p) + count(*,v,p,n2) + count(*,n1,p,n2)}$$

This follows from

$$\begin{aligned} P_3(V|v,n1,p,n2) = {} & \lambda_{v,n1,p} P(V|v,n1,p) \\ & + \lambda_{n1,p,n2} P(V|n1,p,n2) \\ & + \lambda_{v,p,n2} P(V|v,p,n2) \end{aligned}$$

$$P_{mle}(V|v,n1,p) = \frac{count(V,v,n1,p)}{count(*,v,n1,p)}$$

Collins and Brooks' estimation

$$P_3(V|v,n1,p,n2) = \frac{count(V,v,n1,p) + count(V,v,p,n2) + count(V,n1,p,n2)}{count(*,v,n1,p) + count(*,v,p,n2) + count(*,n1,p,n2)}$$

$$P_2(V|v,n1,p,n2) = \frac{count(V,v,p) + count(V,n1,p) + count(V,p,n2)}{count(*,v,p) + count(*,n1,p) + count(*,p,n2)}$$

Collins and Brooks' estimation

$$P_3(V|v,n1,p,n2) = \frac{count(V,v,n1,p) + count(V,v,p,n2) + count(V,n1,p,n2)}{count(*,v,n1,p) + count(*,v,p,n2) + count(*,n1,p,n2)}$$

$$P_2(V|v,n1,p,n2) = \frac{count(V,v,p) + count(V,n1,p) + count(V,p,n2)}{count(*,v,p) + count(*,n1,p) + count(*,p,n2)}$$

$$P_1(V|v,n1,p,n2) = count(V,p) / count(*,p)$$

Collins and Brooks' estimation

$$P_3(V|v,n1,p,n2) = \frac{count(V,v,n1,p) + count(V,v,p,n2) + count(V,n1,p,n2)}{count(*,v,n1,p) + count(*,v,p,n2) + count(*,n1,p,n2)}$$

$$P_2(V|v,n1,p,n2) = \frac{count(V,v,p) + count(V,n1,p) + count(V,p,n2)}{count(*,v,p) + count(*,n1,p) + count(*,p,n2)}$$

$$P_1(V|v,n1,p,n2) = count(V,p) / count(*,p)$$

Combine using **Backoff**

Collins and Brooks' estimation - Back-off

P(V|v,n1,p,n2) =

if count(v,n1,p,n2) > 0

use P_4

else if count(v,n1,p) + count(v,p,n2)+ count(n1,p,n2) > 0

use P_3

else if count(v,p) + count(n1,p)+ count(p,n2, *) > 0

use P_2

else if count(p) > 0

use P_1

else

use P_0 = count(V) / count(V+N)

Collins and Brooks' estimation - Back-off

- Combination of probabilistic model and a heuristic
- Returns a well behaved probability score
  - but not really well motivated by probability theory
- Works well

- → heuristics can be good, if designed well

Collins and Brooks' estimation - Back-off

- Combination of probabilistic model and a heuristic
- Returns a well behaved probability score
  - but not really well motivated by probability theory
- Works well

- → heuristics can be good, if designed well

- **Will be nice to have a method that allows to easily integrate many clues without resorting to heuristics.**

Further improvements

- we've seen
  - (saw,John,with,dog)
- But not
  - (saw,Jack,with,dog)

Can we still say something about the second case?