Simple Probabilistic Modeling & POS

Liad Magen

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 mouse on the hill with a telescope

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 mouse on the hill with a telescope

Ambiguity

Verb NP(1) preposition NP(2)

- 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 mouse on the hill with a telescope

Ambiguity

verb NP(1) preposition NP(2)

verb NP(1) preposition NP(2)

Ambiguity

verb NP(1) preposition NP(2)

I ate pizza with olives

verb NP(1) preposition NP(2)

I ate pizza with friends

Ambiguity

verb NP(1) preposition NP(2)

I ate pizza with olives

verb NP(1) preposition NP(2)

I ate pizza with friends

The N-V PP attachment problem

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

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 mouse on the hill with a telescope

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 mouse on the hill with a telescope

Leaving only the head ("main") words.

- Should we do it?
- Why yes? Why not?

The N-V PP attachment problem

- Given a 4-tuple: (verb, NP1, prep, NP2)
  - talked girl in voice
  - shot elephant in pajamas
  - found sack of money
  - filling garden with flowers
- Predict: V or N , where
  - **V** means a **V-PREP** relation (ate pizza with friends)
  - **N** means a **N-PREP** relation (ate pizza with olives)
- A binary classification task

How do we solve it?

Supervised classification:

- X annotated samples + correct answers (Y):
  - talked girl in voice --> V
  - shot elephant in pajamas --> V
  - found sack of money --> N
  - filling garden with flowers --> V

- Prediction of a new tuple based on previous observation

Steps to solve

1. (always!) Look at the data
2. (always!) Define accuracy measure

   acc = correct / (correct + incorrect)

Conditional Probability

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

Maximum Likelihood Estimation (MLE)

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

Is this reasonable to do?

Data Sparsity; Overfitting

Option #2: Majority baseline

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

Is this reasonable? Would it work?
What score would you expect?

Option #3 – noun1 based

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

Is this reasonable? Would it work?
What score would you expect?

Option #4 – prep based

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

Is this reasonable? Would it work?
What score would you expect?

Option #4 – prep based

P(V | verb, noun1, prep, noun2) ≈ P(V | prep)

Works quite well.

But can we do better?

How about…

P(V| verb, prep) ?

P(V| noun1, prep) ?

P(V| noun1, noun2) ?

P(V| verb, noun1, noun2) ?

P(V| verb, noun1, prep) ?

Or a combination of all?

How do we combine the different probabilities?

- Probability – a review
- A probability function must:
  - Always be positive
  - Sum to one

Combining different probabilities

Obtain a probability through **linear interpolation**:

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

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

Collins and Brooks' estimation

- Interpolate triplets:
  $P_{triplet}$ --> P(V | v, n1 ,p), P(V | v, p, n2), P(V | n1, p, n2)
- Interpolate pairs:
  P_{pair} --> P(V | v, p), P(V | n1, p), P(V | p, n2)

Notice we always include p (the preposition).

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

Why

Collins and Brooks' estimation

- Interpolate triplets:
  $P_{triplet}$ --> **P(V | v, n1 ,p)**, P(V | v, p, n2), P(V | n1, p, n2)

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

- Interpolate pairs:
  P_{pair} --> **P(V | v, p)**, P(V | n1, p), P(V | p, n2)

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

Combining the pair & triplet probabilities

Obtain a probability through **linear interpolation**:

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

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

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

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

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

Collins and Brooks' interpolation:
Gives more weight to frequent training samples.

Collins and Brooks' estimation

- $P_3(V|v,n1,p,n2) =$

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

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' **Back-off Algorithm**

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

if count(v, n1, p, n2) > 0
  return P_4

else if count(v,n1,p) + count(v,p,n2)+ count(n1,p,n2) > 0
  return P_3

else if count(v, p) + count(n1, p)+ count(p, n2, *) > 0
  return P_2

else if count(p) > 0
  return P_1

else:
  return P_0 = count(V) / count(V+N)

Collins and Brooks' **Back-off Algorithm**

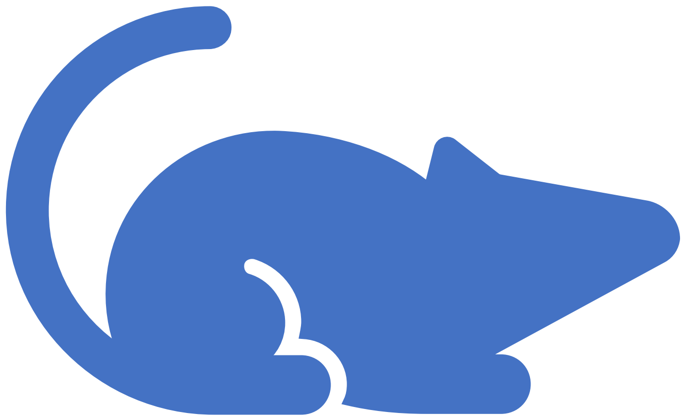
- Combination of probabilistic model and a heuristic
- Returns a well-behaved probability score –
  but not quite well motivated by probability theory
- Works well:

5 Results

The figure below shows the results for the method on the 3097 test sentences, also giving the total count and accuracy at each of the backed-off stages.

Stage	Total Number	Number Correct	Percent Correct
Quadruples	148	134	90.5
Triples	764	688	90.1
Doubles	1965	1625	82.7
Singles	216	155	71.8
Defaults	4	4	100.0
Totals	3097	2606	84.1

[3] At stages 1 and 2 backing off was also continued if $\hat{p}(1|v, n1, p, n2) = 0.5$. ie. the counts were 'neutral' with respect to attachment at this stage.



PP-attachment revisited

We calculated:

$P(V \mid v = \text{saw}, n1 = \text{mouse}, p = \text{with}, n2 = \text{telescope})$

Problems:

- Was not trivial to produce a formula.
- Hard to add more sources of information.

New solution:

- Encode as a binary or multiclass classification.
- Decide on the features.
- Apply a learning algorithm.

PP –attachment as a multiclass classification

Previously, it was defined as a binary classification problem:

Given $x = (v, n1, p, n2)$

Find a $y \in \{V, N\}$

Let's reframe it as a multiclass problem:

$y \in \{ V, N, Other\}$

Our Features:

Single items

- Identity of v
- Identity of p
- Identity of n1
- Identity of n2

Pairs:

- Identity of (v, p)
- Identity of (n1, p)
- Identity of (p, n1)

Triplets:

- Identity of (v, n1, p)
- Identity of (v, p, n2)
- Identity of (n1, p, n2)

Quadruple:

- Identity of (v, n1, p, n2)

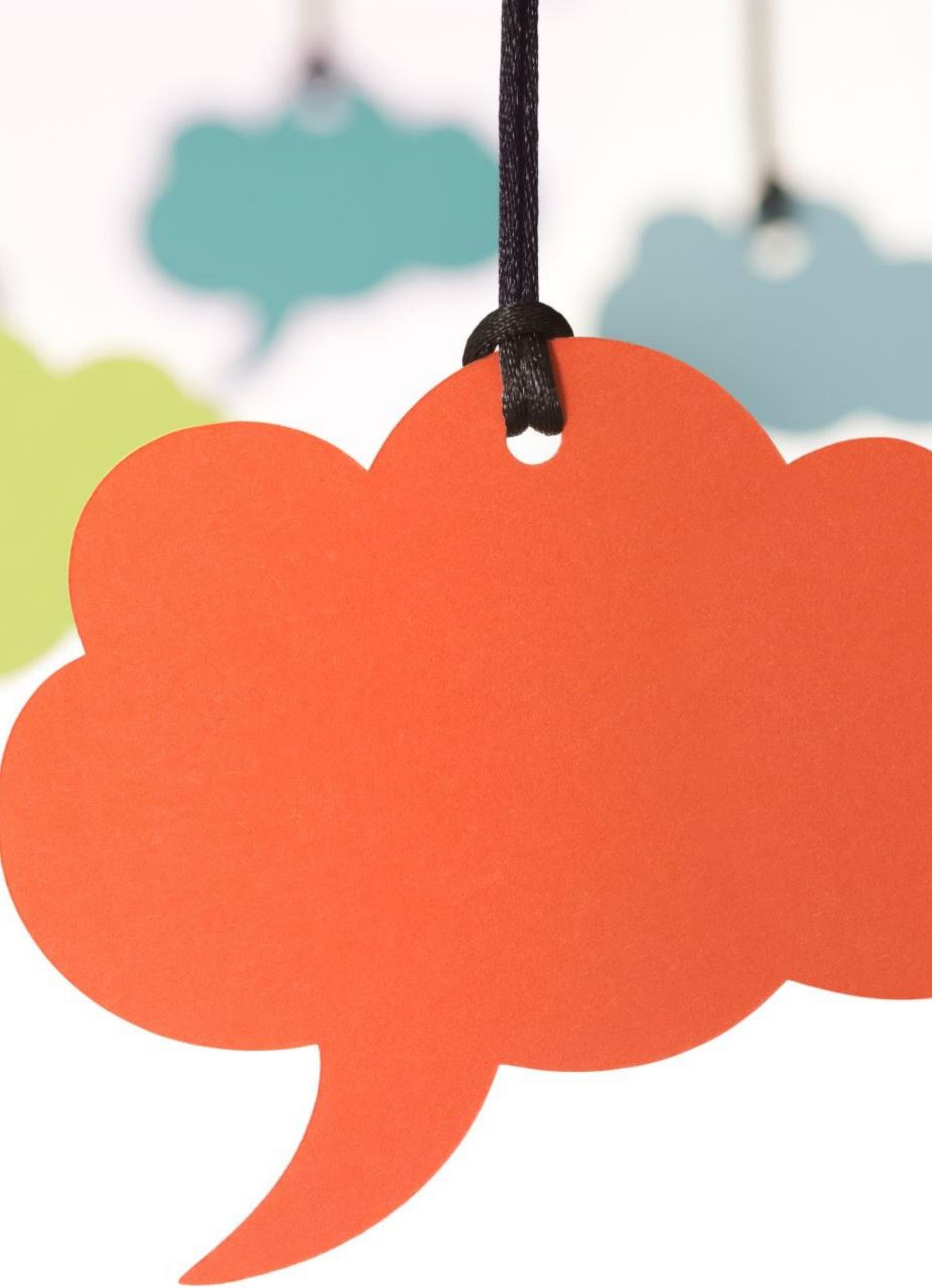
Additional Features

Corpus Level:

- Have we seen the (v, p) pair in a 5-word window in a big corpus?
- Have we seen the (n1, p) pair in a 5-word window in a big corpus?
- Have we seen the (n1, p, n2) triplet in a 5-word window in a big corpus?
    - Also: we can use counts, or binned counts.

Distance:

- Distance (in words) between v and p
- Distance (in words) between n1 and p



...before exercising

- Write a **one sentence** summary of the information we’ve just covered.

- Let's compare your summaries:
  Examine yourself:
  How do your summaries different/similar?

Exercise #4:
Now let's try it out!