ACS Introduction to NLP

Lecture 6: PP Attachment and Lexicalisation

Stephen Clark

Natural Language and Information Processing (NLIP) Group

sc609@cam.ac.uk

PP Attachment

- *See a man with a telescope*
- *Eat the pizza with a fork*
- *Eat the pizza with the anchovies*

- PP attachments are a significant and frequent source of ambiguity
- Resolving PP attachments is difficult, in the worst case requiring world knowledge and general reasoning capabilities
- Tackling the PP attachment problem can give us insights into the more general parse selection problem

Coordination and Relative Clause Attachments (an aside)

- *old boots and shoes*
- *a bank and warehouse guard*

- *Someone shot the servant of the actress who was on the balcony*
- *the lamps near the painting of the house that was damaged in the flood*

- Similar parsing sub-problems can be defined for coordination and relative clause attachments (which are just as hard in the worst case)
- Coordination, in particular, is still a general construction on which statistical parsers perform poorly

[examples on the demo]

PP Attachment and PCFGs

- *Eat the pizza with a fork*
- *Eat the pizza with the anchovies*

- PCFGs only use *structural* probabilities (probabilities of CFG rules)
- Intuitively, the words are only generated at the leaves of the tree, “after” the crucial attachment decision has already been made
- The chosen analyses for the examples above will be the same (either verb attach or noun attach, depending on the rule probabilities)

A Real Example

- Pierre Vinken, 61 years old, *joined the board as a nonexecutive director*
- Did Vinken join as a nonexecutive director? (yes), or is it the board as a nonexecutive director? (no)
- NP-attach (incorrect): (joined ((the board) (as a nonexecutive director)))
- VP-attach (correct): ((joined (the board)) (as a nonexecutive director))

Focus on the 4 Headwords

- Pierre Vinken, 61 years old, *joined the board as a nonexecutive director*
- The decision can be made by considering the 4 headwords: *join, board, as, director*
- Examples like this can be extracted from the Penn Treebank (where 1 indicates noun attach, and 0 verb attach)
- 20,801 tuples for training; 4,039 for development; 3,097 for testing:

0 joined board as director
1 visited house on street
1 is chairman of N.V.
0 bring attention to problem
1 making paper for filters
. . .

[try some on the demo]

A Probabilistic Formulation

$$A_{\text{max}} = \text{arg max}_{A \in \{1,0\}} \hat{P}(A|V = v, N1 = n1, P = p, N2 = n2)$$

where A is the attachment site (1 for noun, 0 for verb), V is the verb, $N1$ is the object of the verb, P is the preposition, $N2$ is the object of preposition

- For example, for *joined the board as a nonexecutive director*:

  $V =$ *joined*, $N1 =$ *board*, $P =$ *as*, $N2 =$ *director*

- Simple algorithm:

  If $\hat{P}(1|$*joined, board, as, director*$) > \hat{P}(0|$*joined, board, as, director*$)$
  attach noun
  else
  attach verb

Baselines and Upper Bounds

Method	Accuracy
Always noun attach	59.0
Most likely based on preposition alone	72.2
Average human (4 head words only)	88.2
Average human (whole sentence)	93.2

from Ratnaparkhi et al. 1994

Probability Estimation

$$\hat{P}(A|v, n1, p, n2) = \frac{f(A, v, n1, p, n2)}{f(v, n1, p, n2)}$$

- But what if $f(v, n1, p, n2) = 0$? (Sound familiar?)
- Maybe we can use similar smoothing methods to what we have already seen for language modelling and tagging

A Backed-Off Model

If $f(v, n1, p, n2) > 0$ then $\hat{P}(A|v, n1, p, n2) = \frac{f(A,v,n1,p,n2)}{f(v,n1,p,n2)}$

else if $f(\textit{triples}) > 0$ then $\hat{P}(A|v, n1, p, n2) = \frac{f(A,\textit{triples})}{f(\textit{triples})}$

else if $f(\textit{pairs}) > 0$ then $\hat{P}(A|v, n1, p, n2) = \frac{f(A,\textit{pairs})}{f(\textit{pairs})}$

else if $f(\textit{singleton}) > 0$ then $\hat{P}(A|v, n1, p, n2) = \frac{f(A,\textit{singleton})}{f(\textit{singleton})}$

- How should we break the 4-tuple into triples, pairs and singletons?
- The preposition is crucial, so always include that

A Backed-Off Model

If $f(v, n1, p, n2) > 0$ then $\hat{P}(A|v, n1, p, n2) = \frac{f(A,v,n1,p,n2)}{f(v,n1,p,n2)}$

else if $f(v, n1, p) + f(v, p, n2) + f(n1, p, n2) > 0$ then

$\hat{P}(A|v, n1, p, n2) = \frac{f(A,v,n1,p)+f(A,v,p,n2)+f(A,n1,p,n2)}{f(v,n1,p)+f(v,p,n2)+f(n1,p,n2)}$

else if $f(v, p) + f(n1, p) + f(p, n2) > 0$ then

$\hat{P}(A|v, n1, p, n2) = \frac{f(A,v,p)+f(A,n1,p)+f(A,p,n2)}{f(v,p)+f(n1,p)+f(p,n2)}$

else if $f(p) > 0$ then

$\hat{P}(A|v, n1, p, n2) = \frac{f(A,p)}{f(p)}$

else $\hat{P}(1|v, n1, p, n2) = 1.0$ (i.e. default to noun attach)

Results

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

Discussion

- Perhaps surpisingly, using a higher value for the cutoff parameter reduced performance
- In other words, for this task, it is always better to use a non-zero count rather than back-off to a more general context (even if the count is 1)
- Morphological analysis (replacing years with YEAR, numbers with NUM, proper names with NAME, words with lemmas) gives only a small improvement (0.4%)

Other Approaches

- Some work on unsupervised PP attachment
  - based on the idea that some attachments are unambiguous; eg *the pizza with anchovies was tasty*, *eating with a fork is usual in the UK*
- Lots of work on using semantic information, e.g. from WordNet (if I know anchovies are a kind of food, fork is an implement . . .)

References

- Michael Collins and James Brooks (1995), Prepositional Phrase Attachment through a Backed-off Model. Proceedings of the Third Workshop on Very Large Corpora.
- Ratnaparkhi, Reynar and Roukos (1994). A Maximum Entropy Model for Prepositional Phrase Attachment. Proceedings of the ARPA Workshop on Human Language Technology
- Ratnaparkhi (1998), Statistical Models for Unsupervised Prepositional Phrase Attachment, ACL-COLING 1998
- Stetina and Nagoa (1997), Corpus based PP attachment ambiguity resolution with a semantic dictionary, Workshop on Very Large Corpora
- For more recent papers, do a search with *PP attachment ambiguity*