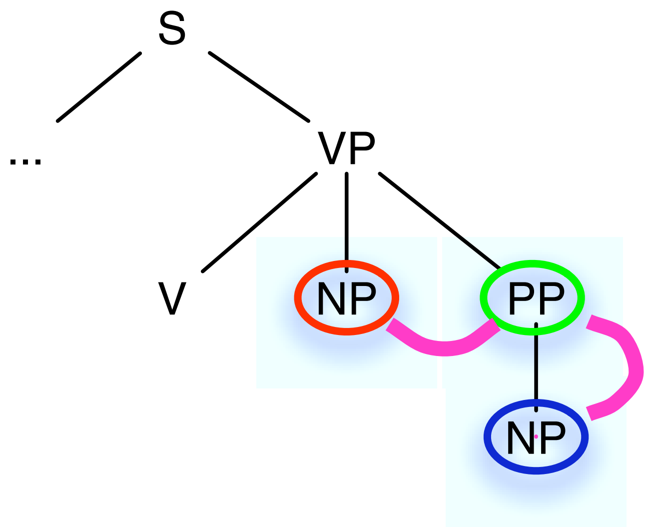
Semantic Role Labelling with Tree Conditional Random Fields

Trevor Cohn and Philip Blunsom

University of Melbourne

June 2005

Motivation



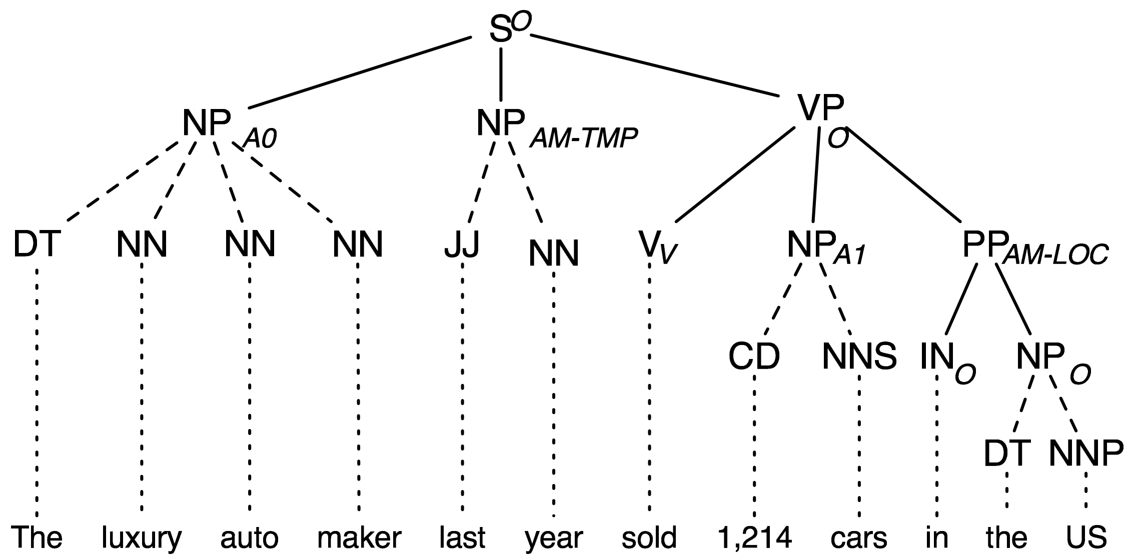
- Independent per-node classification makes common errors:
  - child nodes of arguments should not be flagged as arguments
  - arguments should be predicted once or not at all

Approach

- Used CRFs for SRL tagging task
  - label each constituent in parse tree
  - use parse tree as model of adjacency (random field)
  - cleaner solution than per-token tagging or per-node classification
  - labelling constraints enforcable over parent-child nodes

- Efficient CRF inference methods generalise to trees
  - tractable in the absense of loops
  - application of Pearl's $\{$sum,max$\}$-product algorithm

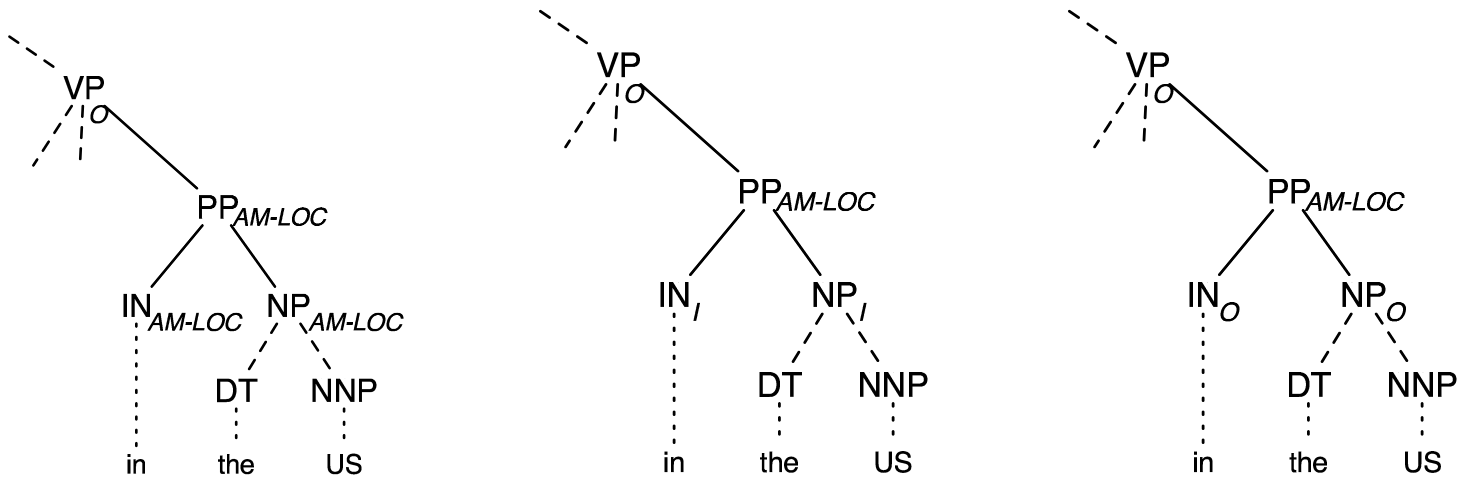
Graphical structure



- Cliques for each single node and each parent and child pair
- Pruning used to limit tree size

Labelling method

- Tried three different labelling strategies



argument　　inside　　outside

- Outside best performing

CRFs

- Generic CRF definition:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp \sum_{c \in \mathcal{C}} \sum_{k} \lambda_k f_k(c, \mathbf{y}_c, \mathbf{x})$$

- Training involves optimising objective (eg. log-likelihood)

$$E_{\tilde{p}(\mathbf{x},\mathbf{y})}[f_k] - E_{p(\mathbf{y}|\mathbf{x})}[f_k] = 0$$

  - use sum-product to calculate marginals needed by $E_{p(\mathbf{y}|\mathbf{x})}[f_k]$
  - use max-product to find best labelling

Results

- Similar features to Xue & Palmer (2004), Pradhan et al. (2005)
  - using Collins parses

- F_1 scores of 71.17 (dev), 73.10 (test WSJ) and 63.63 (test Brown)

- Findings:
  - CRF improved over maxent classifier (+1%)
  - pruning was detrimental to generalisation performance (-1%)
  - Charniak parses more useful (+3%) [recent finding]
  - Very few inconsistent ancestor/dependent labellings
  - Quite a number of duplicate argument predictions