MST-based Dependency Parsing: Learning

Pawan Goyal

CSE, IIT Kharagpur

Week 6, Lecture 5

Arc weights as linear classifiers

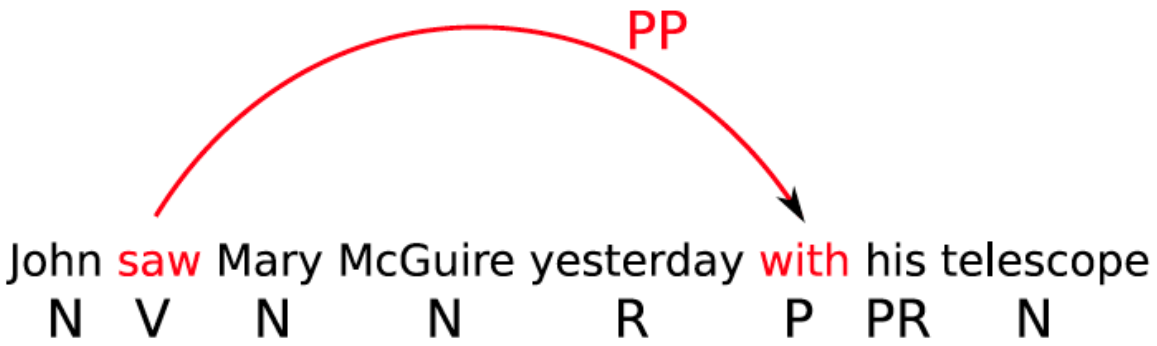
$$w_{ij}{}^{k} = w.f(i,j,k)$$

Arc weights as linear classifiers

$$w_{ij}{}^{k} = w.f(i,j,k)$$

- Arc weights are a linear combination of features of the arc $f(i,j,k)$ and a corresponding weight vector w
- What arc features?

Arc Features $f(i,j,k)$

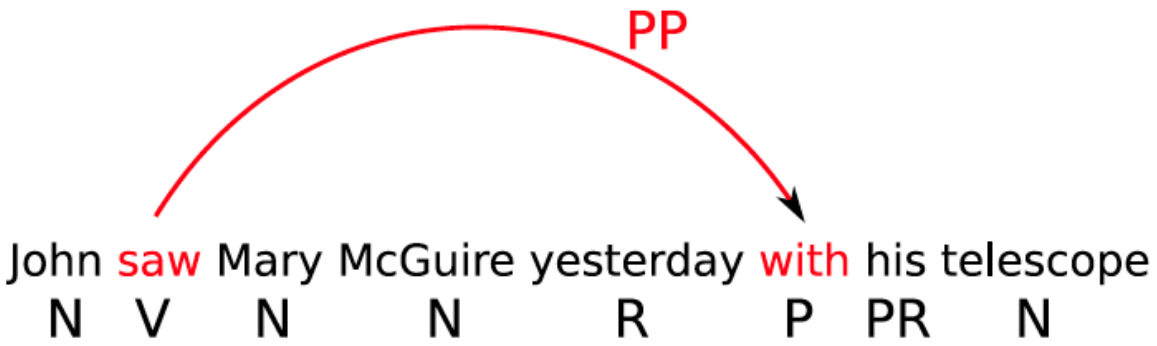


Features

Identities of the words w_i and w_j for a label l_k

head = saw & dependent=with

Arc Features $f(i,j,k)$

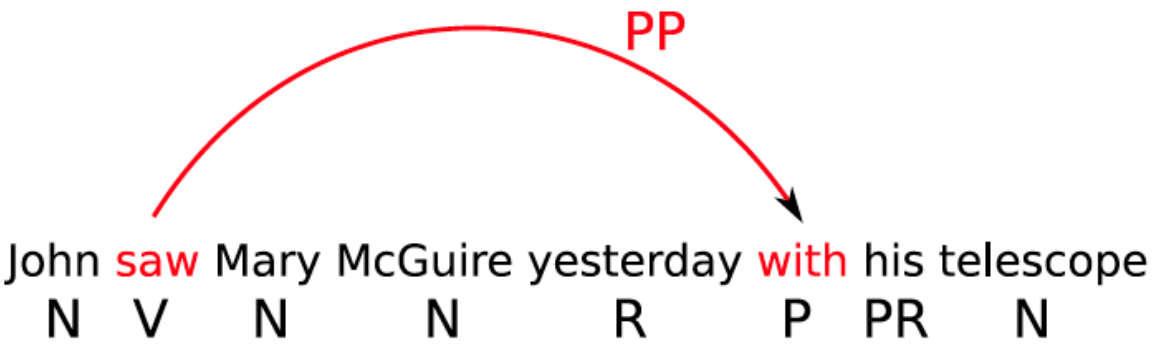


Features

Part-of-speech tags of the words w_i and w_j for a label l_k

head-pos = Verb & dependent-pos=Preposition

Arc Features $f(i,j,k)$



Features

Part-of-speech of words surrounding and between w_i and w_j

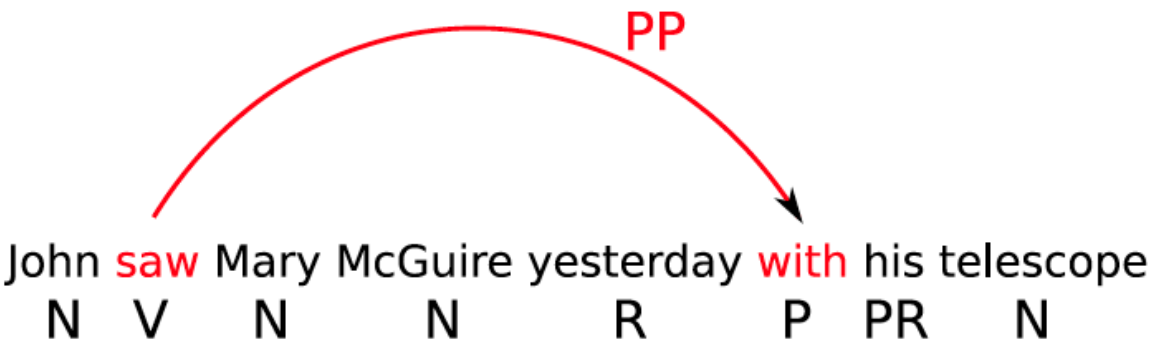
inbetween-pos = Noun

inbetween-pos = Adverb

dependent-pos-right = Pronoun

head-pos-left=Noun

Arc Features $f(i,j,k)$



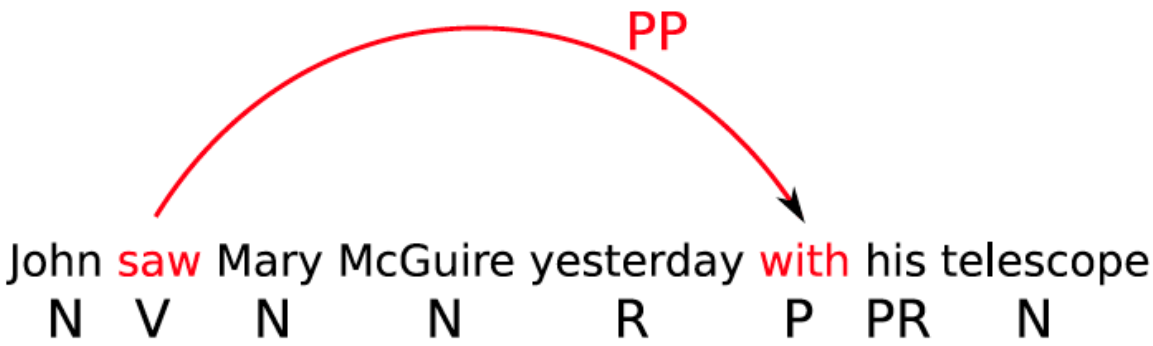
Features

Number of words between w_i and w_j, and their orientation

arc-distance = 3

arc-direction = right

Arc Features $f(i,j,k)$



Features

- Combinations
  head-pos=Verb & dependent-pos=Preposition & arc-label=PP
  head-pos=Verb & dependent=with & arc-distance=3
- No limit : any feature over arc (i,j,k) or input x

Learning the parameters

- Re-write the inference problem

$$G = \arg\max_{G \in T(G_x)} \sum_{(i,j,k) \in G} {w_{ij}}^k$$

$$= \arg\max_{G \in T(G_x)} w \cdot \sum_{(i,j,k) \in G} f(i,j,k)$$

$$= \arg\max_{G \in T(G_x)} w \cdot f(G)$$

Learning the parameters

- Re-write the inference problem

$$G = \arg\max_{G \in T(G_x)} \sum_{(i,j,k) \in G} {w_{ij}}^k$$

$$= \arg\max_{G \in T(G_x)} w \cdot \sum_{(i,j,k) \in G} f(i,j,k)$$

$$= \arg\max_{G \in T(G_x)} w \cdot f(G)$$

- Which can be plugged into learning algorithms

Inference-based Learning

Training data: $T = \{(x_t, G_t)\}_{t=1}^{|T|}$

1. $w^{(0)} = 0; i = 0$
2. for $n : 1..N$
3. for $t : 1..|T|$
4. Let $G' = argmax_{G'} w^{(i)}.f(G')$
5. if $G' \neq G_t$
6. $w^{(i+1)} = w^{(i)} + f(G_t) - f(G')$
7. $i = i + 1$
8. return w^i

Example

Suppose you are training MST Parser for dependency and the sentence, "John saw Mary" occurs in the training set. Also, for simplicity, assume that there is only one dependency relation, "rel". Thus, for every arc from word w_i to w_j, your features may be simplified to depend only on words w_i and w_j and not on the relation label.
Below is the set of features

- f_1: pos(w_i) = Noun and pos(w_j) = Noun
- f_2: pos(w_i) = Verb and pos(w_j) = Noun
- f_3: w_i = Root and pos(w_j) = Verb
- f_4: w_i = Root and pos(w_j) = Noun
- f_5: w_i = Root and w_j occurs at the end of sentence
- f_6: w_i occurs before w_j in the sentence
- f_7: pos(w_i) = Noun and pos(w_j) = Verb

The feature weights before the start of the iteration are: $\{3, 20, 15, 12, 1, 10, 20\}$. Determine the weights after an iteration over this example.