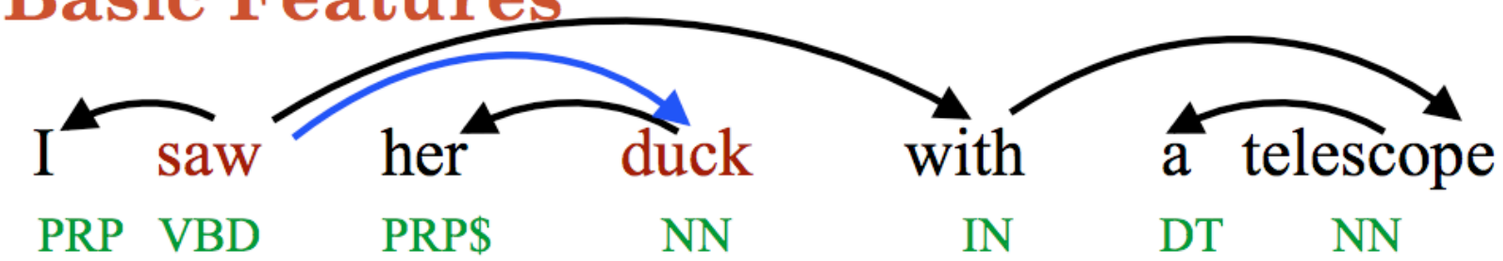
Introduction to Syntax and Parsing
ACS 2015/16
Stephen Clark

L4: The Perceptron Parsing Model

UNIVERSITY OF
CAMBRIDGE

Edge-Based Linear Model

Basic Features



- Uni-gram features
- Bi-gram features
- In between POS features
- Surrounding word POS features

Saw_VBD, saw, VBD duck_NN, duck, NN
saw_VBD_duck_NN, VBD_duck_NN, saw_duck_NN, saw_VBD_NN, saw_VBD_duck, Saw_duck, VBD_NN
VBD_PRP$_NN
VBD_PRP$_PRP$_NN, PRP_VBD_PRP$_NN, VBD_PRP$_NN_IN, PRP_VBD_NN_IN

taken from Wang and Zhang, NAACL tutorial 2010

$$score(x_i \rightarrow x_j) = \sum_k \lambda_k \, . \, f_k(x_i \rightarrow x_j)$$

Features in the MST Parser

a)

Basic Uni-gram Features
p-word, p-pos
p-word
p-pos
c-word, c-pos
c-word
c-pos

b)

Basic Bi-gram Features
p-word, p-pos, c-word, c-pos
p-pos, c-word, c-pos
p-word, c-word, c-pos
p-word, p-pos, c-pos
p-word, p-pos, c-word
p-word, c-word
p-pos, c-pos

c)

In Between POS Features
p-pos, b-pos, c-pos
Surrounding Word POS Features
p-pos, p-pos+1, c-pos-1, c-pos
p-pos-1, p-pos, c-pos-1, c-pos
p-pos, p-pos+1, c-pos, c-pos+1
p-pos-1, p-pos, c-pos, c-pos+1

Table 1: Features used by system. p-word: word of parent in dependency edge. c-word: word of child. p-pos: POS of parent. c-pos: POS of child. p-pos+1: POS to the right of parent in sentence. p-pos-1: POS to the left of parent. c-pos+1: POS to the right of child. c-pos-1: POS to the left of child. b-pos: POS of a word in between parent and child.

taken from McDonald et al.

Global Linear Model

$$Score(\tau) = \sum_{x_i \to x_j \in \tau} \sum_k w_k \,.\, f_k(x_i \to x_j)$$

$$= \sum_k w_k \,.\, \sum_{x_i \to x_j \in \tau} f_k(x_i \to x_j)$$

$$= \sum_k w_k \,.\, f_k(\tau)$$

$$= \mathbf{w}.\mathbf{F}(\tau)$$

Generic Online Learning

Training data: $\mathcal{T} = \{(\boldsymbol{x}_t, \boldsymbol{y}_t)\}_{t=1}^{T}$

1. $\mathbf{w}_0 = 0;\ \mathbf{v} = 0;\ i = 0$
2. for n : 1..N
3. for t : 1..T
4. $\mathbf{w}^{(i+1)}$ = update $\mathbf{w}^{(i)}$ according to instance $(\boldsymbol{x}_t, \boldsymbol{y}_t)$
5. $\mathbf{v} = \mathbf{v} + \mathbf{w}^{(i+1)}$
6. $i = i + 1$
7. $\mathbf{w} = \mathbf{v}/(N * T)$

taken from McDonald et al.

The Perceptron Update

Given sentence $\mathbf{x}_t$ and correct tree $\mathbf{y}_t$

$$\mathbf{z}_t = \arg\max_{\mathbf{z}} \mathbf{w}_{t-1} \,.\, \mathbf{F}(\mathbf{x}_t, \mathbf{z})$$

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \mathbf{F}(\mathbf{x}_t, \mathbf{y}_t) - \mathbf{F}(\mathbf{x}_t, \mathbf{z}_t)$$

UNIVERSITY OF CAMBRIDGE

CoNLL Shared Task Data

Multilingual										
	Ar	Ba	Ca	Ch	Cz	En	Gr	Hu	It	Tu
Language family	Sem.	Isol.	Rom.	Sin.	Sla.	Ger.	Hel.	F.-U.	Rom.	Tur.
Annotation	d	d	c+f	c+f	d	c+f	d	c+f	c+f	d
Training data										
Tokens (k)	112	51	431	337	432	447	65	132	71	65
Sentences (k)	2.9	3.2	15.0	57.0	25.4	18.6	2.7	6.0	3.1	5.6
Tokens/sentence	38.3	15.8	28.8	5.9	17.0	24.0	24.2	21.8	22.9	11.6
LEMMA	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
No. CPOSTAG	15	25	17	13	12	31	18	16	14	14
No. POSTAG	21	64	54	294	59	45	38	43	28	31
No. FEATS	21	359	33	0	71	0	31	50	21	78
No. DEPREL	29	35	42	69	46	20	46	49	22	25
No. DEPREL H=0	18	17	1	1	8	1	22	1	1	1
% HEAD=0	8.7	9.7	3.5	16.9	11.6	4.2	8.3	4.6	5.4	12.8
% HEAD left	79.2	44.5	60.0	24.7	46.9	49.0	44.8	27.4	65.0	3.8
% HEAD right	12.1	45.8	36.5	58.4	41.5	46.9	46.9	68.0	29.6	83.4
HEAD=0/sentence	3.3	1.5	1.0	1.0	2.0	1.0	2.0	1.0	1.2	1.5
% Non-proj. arcs	0.4	2.9	0.1	0.0	1.9	0.3	1.1	2.9	0.5	5.5
% Non-proj. sent.	10.1	26.2	2.9	0.0	23.2	6.7	20.3	26.4	7.4	33.3
Punc. attached	S	S	A	S	S	A	S	A	A	S
DEPRELS for punc.	10	13	6	29	16	13	15	1	10	12

Taken from Nivre et al. (2007)

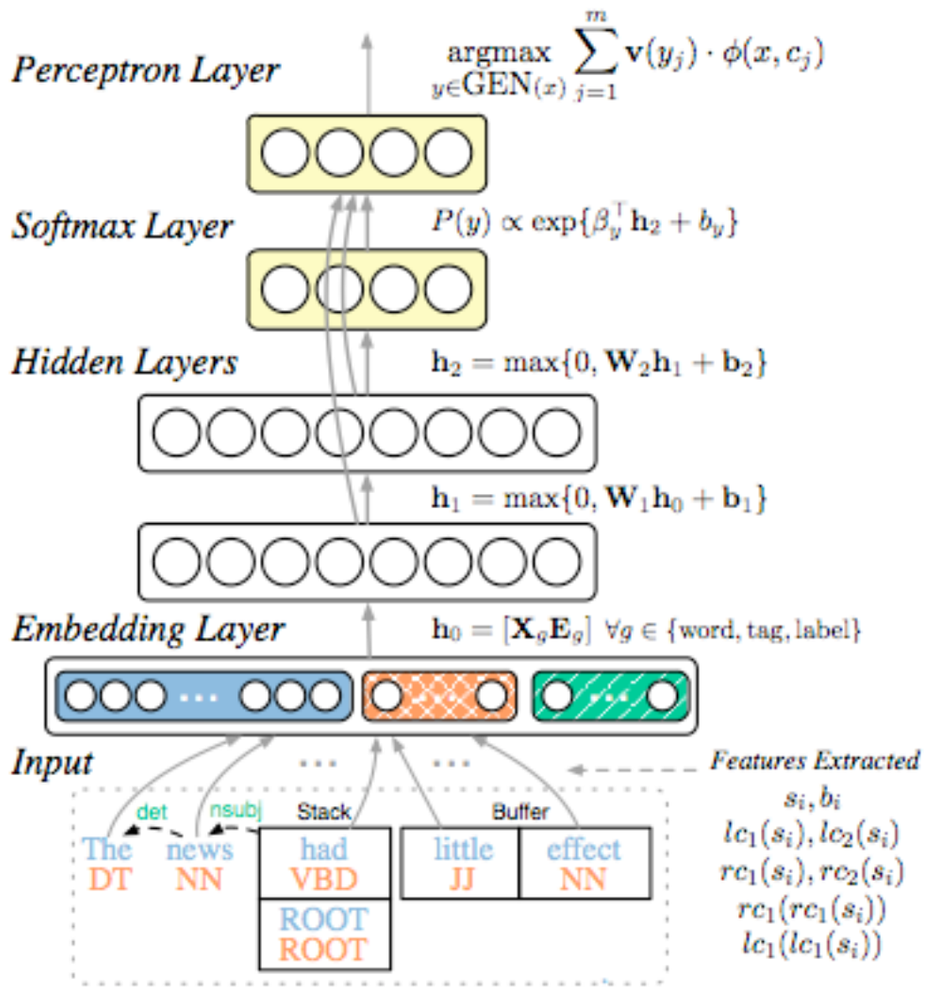
Graph-based vs. Transition-based

Table 1
Labeled parsing accuracy for top-scoring systems at CoNLL-X (Buchholz and Marsi 2006).

Language	Graph-based (McDonald, Lerman, and Pereira 2006)	Transition-based (Nivre et al. 2006)
Arabic	66.91	66.71
Bulgarian	87.57	87.41
Chinese	85.90	86.92
Czech	80.18	78.42
Danish	84.79	84.77
Dutch	79.19	78.59
German	87.34	85.82
Japanese	90.71	91.65
Portuguese	86.82	87.60
Slovene	73.44	70.30
Spanish	82.25	81.29
Swedish	82.55	84.58
Turkish	63.19	65.68
Average	80.83	80.75

Taken from McDonald & Nivre (2011)

State-of-the-Art (2015)



Taken from Weiss et al. (2015)

UNIVERSITY OF CAMBRIDGE

Accuracy League Table (2015)

Method	UAS	LAS	Beam
Graph-based			
Bohnet (2010)	92.88	90.71	n/a
Martins et al. (2013)	92.89	90.55	n/a
Zhang and McDonald (2014)	93.22	91.02	n/a
Transition-based			
*Zhang and Nivre (2011)	93.00	90.95	32
Bohnet and Kuhn (2012)	93.27	91.19	40
Chen and Manning (2014)	91.80	89.60	1
S-LSTM (Dyer et al., 2015)	93.20	90.90	1
Our Greedy	93.19	91.18	1
Our Perceptron	**93.99**	**92.05**	8

Results for English WSJ

Taken from Weiss et al. (2015)