Incremental Parsing with Minimal Features Using Bi-Directional LSTM

James Cross
EECS, Oregon State University
crossj@oregonstate.edu

Liang Huang
EECS, Oregon State University
liang.huang.sh@gmail.com

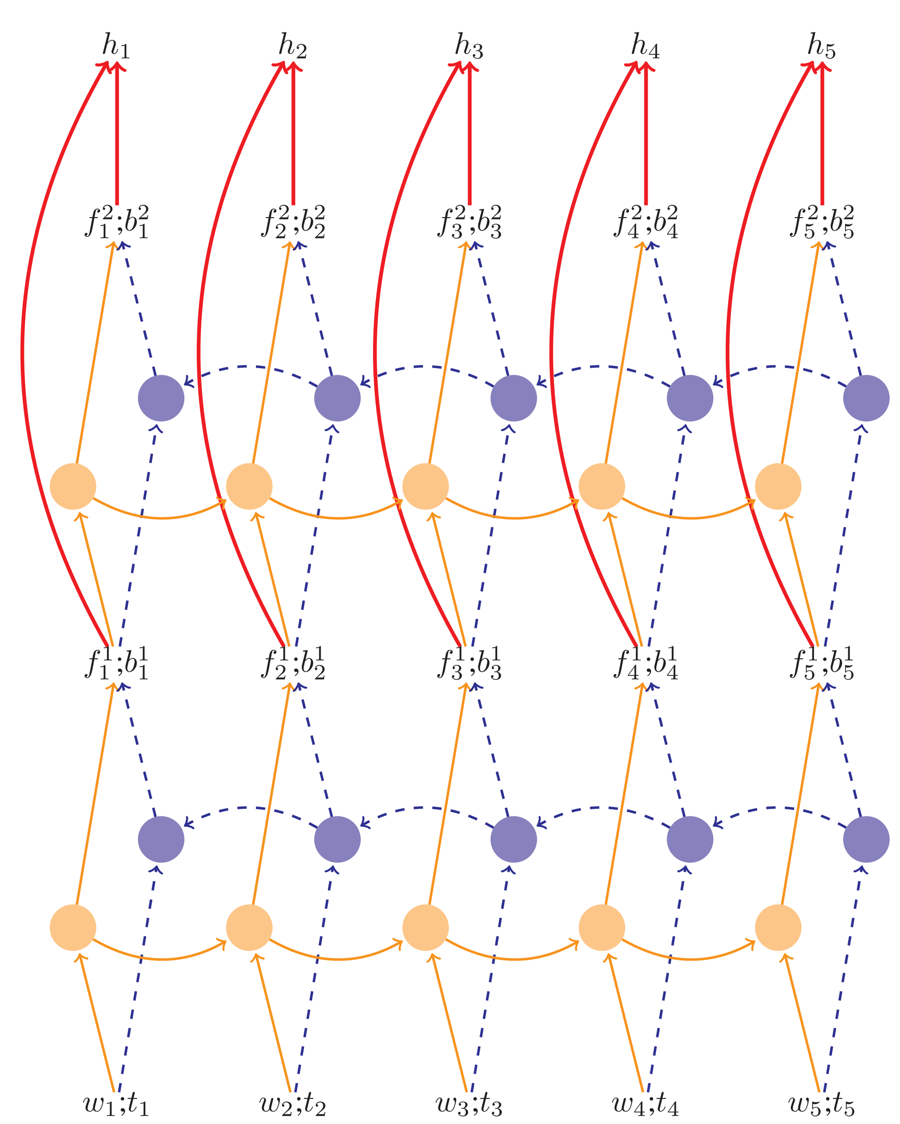
Oregon State UNIVERSITY

Overview

- We use bi-directional LSTM to train greedy transition parsers with a bare minimum of features.
- A new transition system for constituency parsing offers competitive performance even with greedy inference.
- State-of-the-art performance among greedy parsers (at time of submission) for both dependency and constituency parsers.

LSTM Position Features

Sentences are modeled with a recurrent neural network using **word** and **part-of-speech embeddings** (learned only from the training data) as input.



We found the best results by concatenating the output of each of two subsequent bi-directional LSTM layers.

Parser	UAS	LAS
One-layer Bi-LSTM†	93.31	91.01
† - Backward-LSTM	91.12	88.72
† - Forward-LSTM	91.85	88.39
† - tag embeddings	92.46	89.81

Ablation studies on PTB dev set (wsj 22). Forward and backward context, and part-of-speech input were all critical to strong performace.

Deductive Systems

Arc-Standard (Dependency)

input: $w_0 \dots w_{n-1}$

axiom $\langle \epsilon,\ 0\rangle : \emptyset$

shift $$\frac{\langle S,\ j\rangle : A}{\langle S|j,\ j+1\rangle : A}\ j<n$$

re$_\curvearrowleft$ $$\frac{\langle S|s_1|s_0,\ j\rangle : A}{\langle S|s_0,\ j\rangle : A \cup \{s_1 \curvearrowleft s_0\}}$$

goal $\langle s_0,\ n\rangle : A$

Shift-Promote-Adjoin (Constituency)

shift $$\frac{\langle S,\ j\rangle}{\langle S \mid j,\ j+1\rangle}\ j<n$$

pro(X) $$\frac{\langle S \mid t,\ j\rangle}{\langle S \mid X(t),\ j\rangle}$$

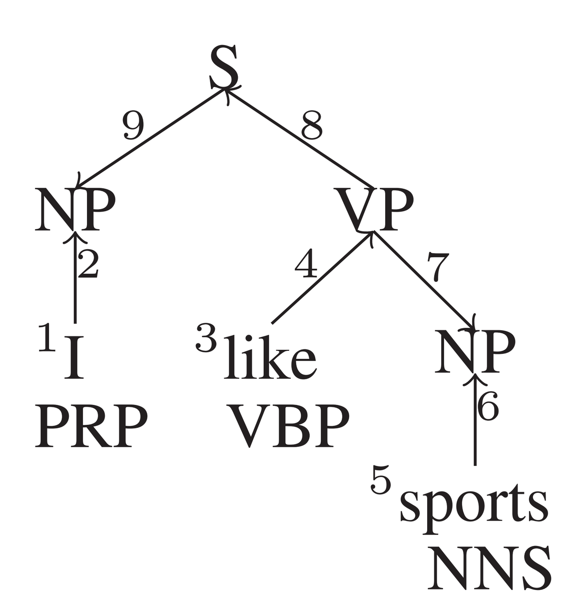
adj$_\curvearrowleft$ $$\frac{\langle S \mid t \mid X(t_1...t_k),\ j\rangle}{\langle S \mid X(t, t_1...t_k),\ j\rangle}$$

goal $\langle s_0,\ n\rangle$

Shift-Promote-Adjoin Constituency Parsing

We propose a novel tranistion system for constituency parsing, inpsired by arc-standard dependency parsing, which:

- Does not require binarization.
- Has only $3 + X$ actions, where X is the number of non-terminal labels.
- Promotes the head constituent of each phrase before attaching siblings.



1shift (I)	6pro (NP)
2pro (NP)	7adj$_\curvearrowleft$
3shift (like)	8pro (S)
4pro (VP)	9adj$_\curvearrowleft$
5shift (sports)	

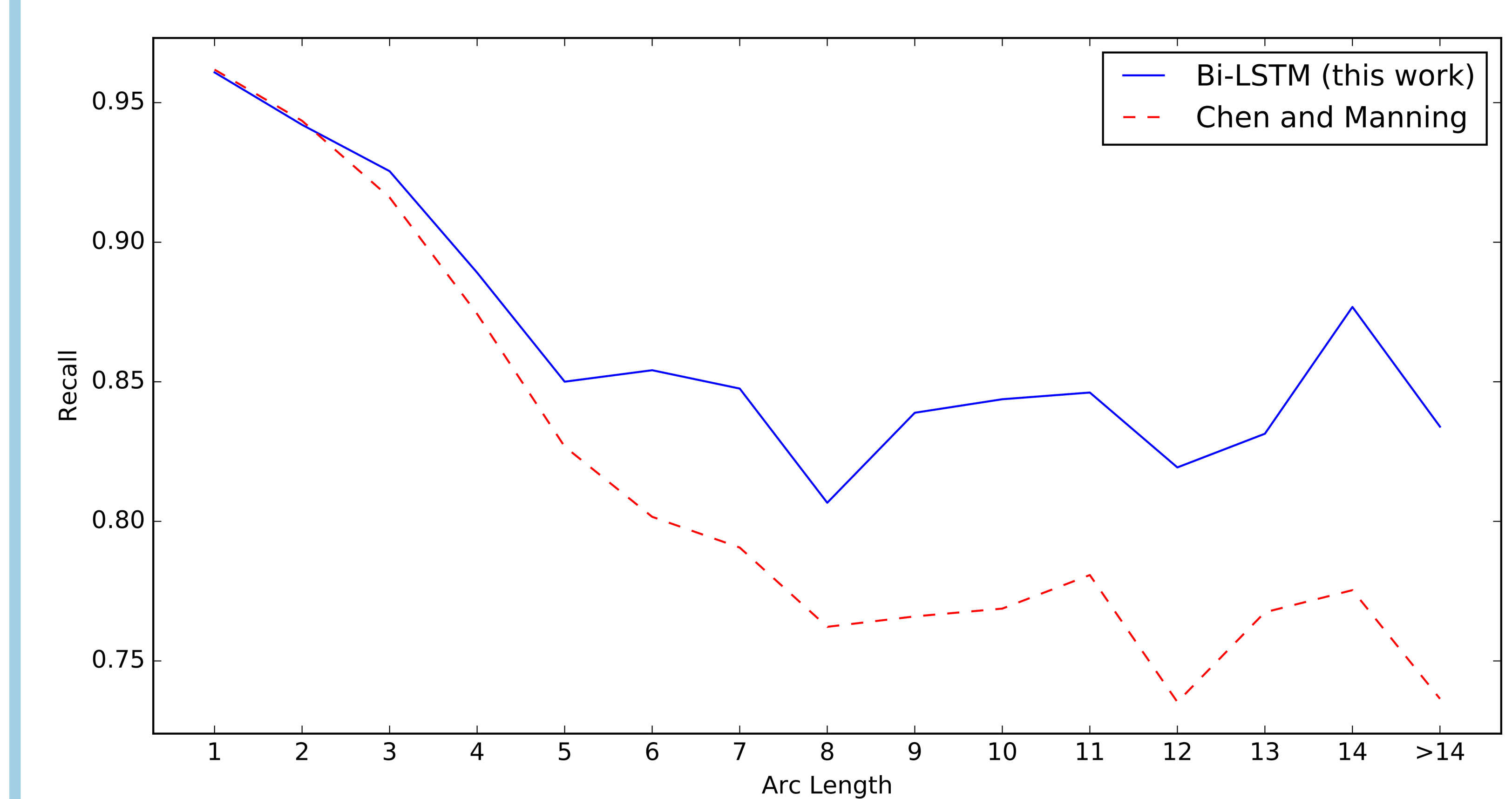
Features

	dependency	constituency
positional	s_1, s_0, q_0	s_1, s_0, q_0, s_1.left, s_0.left
labels	-	s_0.{left, right, root, head} s_1.{left, right, root, head}

Experimental Results

Dependency	English UAS	English LAS	Chinese UAS	Chinese LAS
Chen and Manning 2014	91.8	89.6	83.9	82.4
Dyer et al. 2015	93.1	90.9	**87.2**	85.7
Bi-LSTM	93.21	91.16	85.53	84.89
2-Layer Bi-LSTM	**93.42**	91.36	86.35	85.71

Greedy-parser accuracy on English and Chinese Penn Treebank test sets. Current state of the art (94.61 English UAS) by Andor et al. Globally Normalized Transition-Based Neural Networks. *ACL* 2016.



Recall on dependency arcs of various lengths in PTB dev set. The Bi-LSTM parser is particularly good at predicting longer arcs.

Constituency	English greedy	English beam	Chinese greedy	Chinese beam
Zhu et al. (2013)	86.08	90.4	75.99	85.6
Mi & Huang (2005)	84.95	90.8	75.61	83.9
Bi-LSTM	89.75	-	79.44	-
2-Layer Bi-LSTM	**89.95**	-	**80.13**	-

F_1 score on English and Chinese Penn Treebank test sets for transition-based constituency parsers. In our upcoming paper (James Cross and Liang Huang. Span-Based Constituency Parsing with a Structure-Label System and Dynamic Oracles. *EMNLP* 2016 (to appear)), we describe a system with state-of-the-art accuracy (91.4 for English) with greedy inference.