Structured Training for Neural Network Transition-Based Parsing

David Weiss, Chris Alberti, Michael Collins, Slav Petrov

Presented by: Shayne Longpre

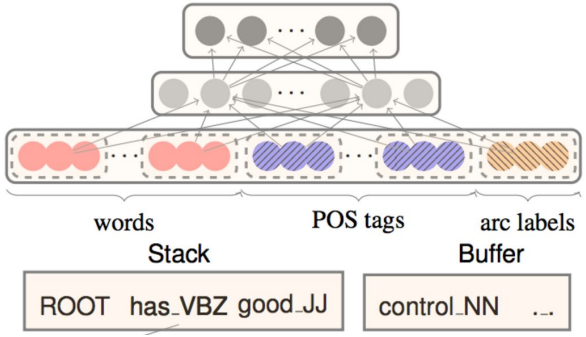
What is SyntaxNet?

- 2016/5: Google announces the “World’s Most Accurate Parser Goes Open Source”
- SyntaxNet (2016): New, fast, performant Tensorflow framework for syntactic parsing.
- Now supports 40 languages -- Parsey McParseface’s 40 ‘cousins’

What is SyntaxNet?

- 2016/5: Google announces the “World’s Most Accurate Parser Goes Open Source”
- SyntaxNet (2016): New, fast, performant Tensorflow framework for syntactic parsing.
- Now supports 40 languages -- Parsey McParseface’s 40 ‘cousins’

Chen & Manning (2014) → Weiss et al. (2015) → Andor et al. (2016)



\+ Unlabelled Data
\+ Tune Model
\+ Structured Perceptron & Beam Search

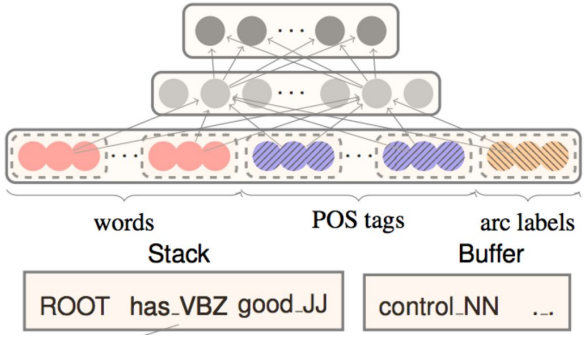
\+ Global Normalization

= **SyntaxNet**

What is SyntaxNet?

- 2016/5: Google announces the “World’s Most Accurate Parser Goes Open Source”
- SyntaxNet (2016): New, fast, performant Tensorflow framework for syntactic parsing.
- Now supports 40 languages -- Parsey McParseface’s 40 ‘cousins’

Chen & Manning (2014)



Weiss et al. (2015)

+ Unlabelled Data
+ Tune Model
+ Structured Perceptron & Beam Search

Andor et al. (2016)

+ Global Normalization

= **SyntaxNet**

3 New Contributions (… since Q2 in Assignment 2)

1. Leverage Unlabelled Data -- “Tri-Training”
2. Tuned Neural Network Model
3. Final Layer: Structured Perceptron w/ Beam Search

1. Tri-Training: *Leverage Unlabelled Data*

High Performance Parser (A)

High Performance Parser (B)

Agree on dependency parse

Disagree on dependency parse

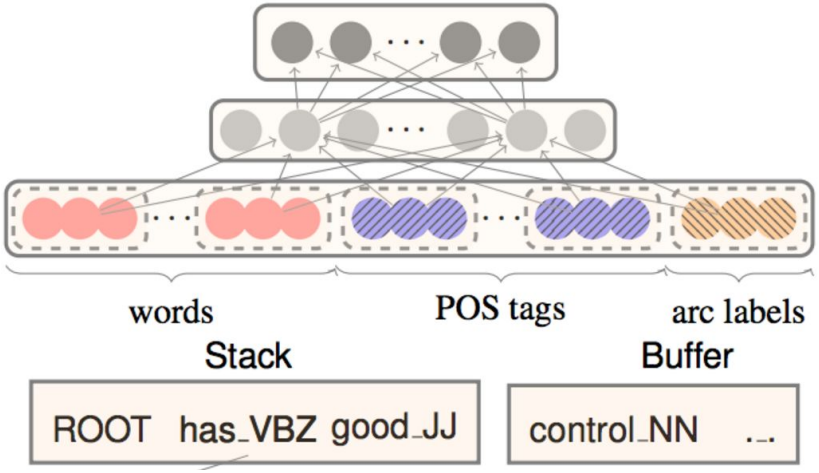
Unlabelled Data

"Tri-Training" (Li et al, 2014)

Labelled Data

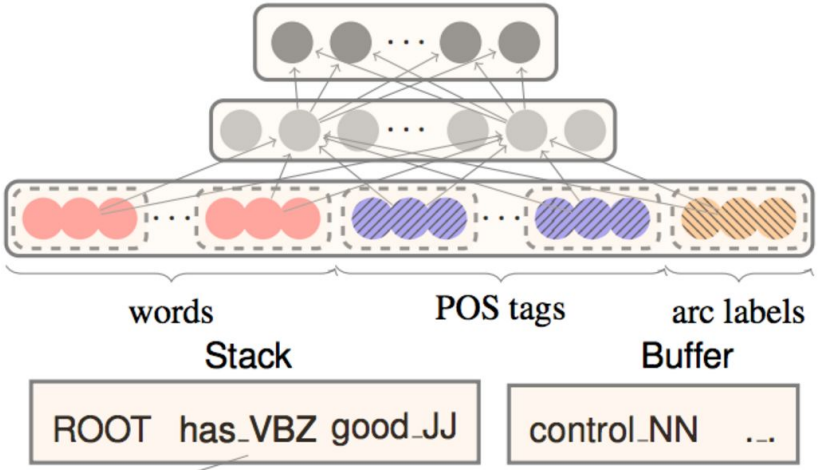
2. Model Changes

Chen & Manning (2014):

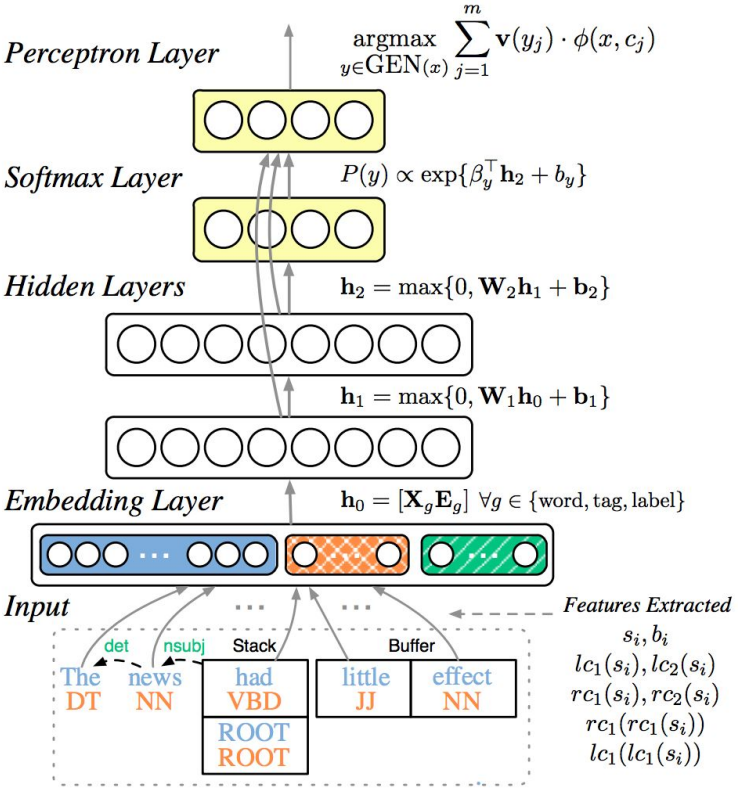


2. Model Changes

Chen & Manning (2014):

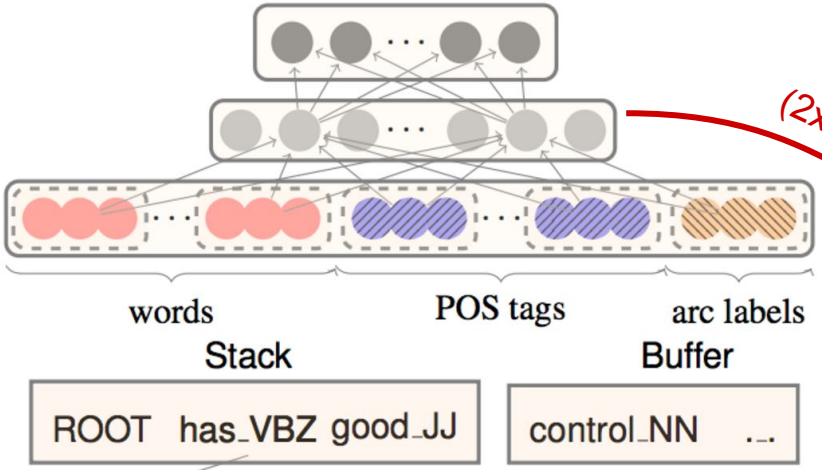


Weiss et al. (2015):



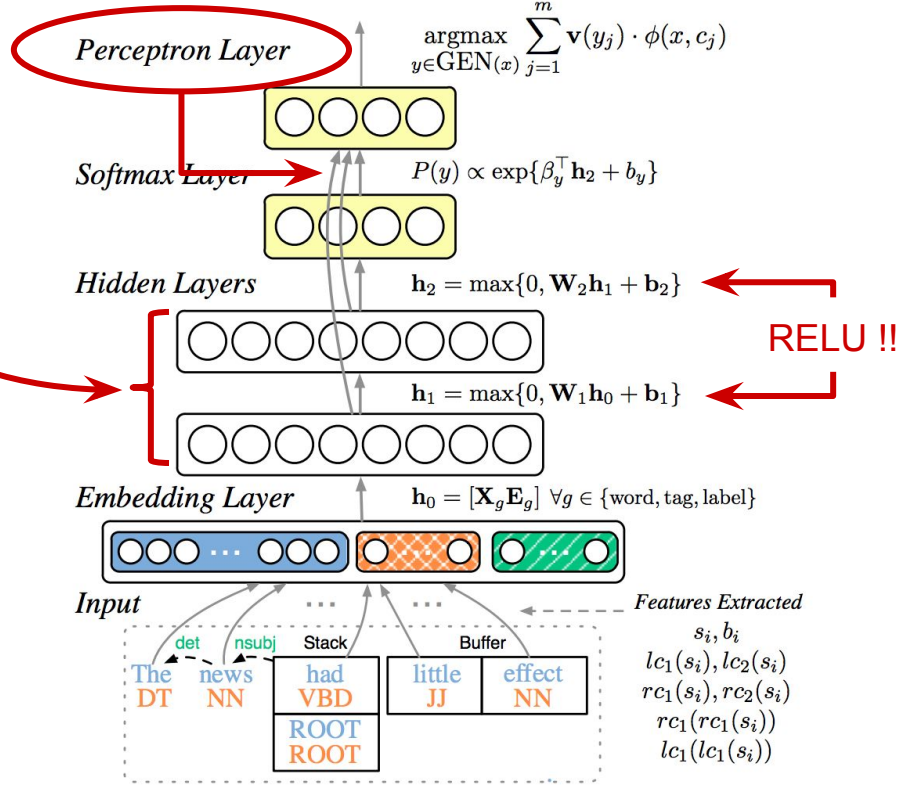
2. Model Changes

Chen & Manning (2014):



(2x hidden)

Weiss et al. (2015):



3. Structured Perceptron Training + Beam Search

Problem: Greedy algorithms are unable to look beyond one step ahead, or recover from incorrect decisions.

3. Structured Perceptron Training + Beam Search

Problem: Greedy algorithms are unable to look beyond one step ahead, or recover from incorrect decisions.

Solution: Look forward -- search the tree of possible transition sequences.

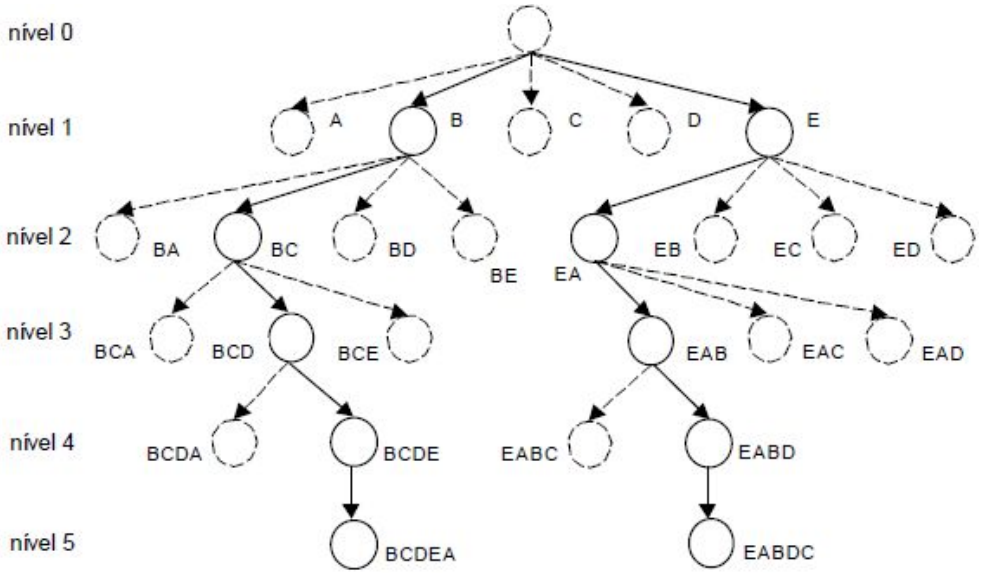


Figura 1 – Árvore de busca utilizando o *beam search*

3. Structured Perceptron Training + Beam Search

Problem: Greedy algorithms are unable to look beyond one step ahead, or recover from incorrect decisions.

Solution: Look forward -- search the tree of possible transition sequences.

- Keep track of *K* top partial transition sequences up to depth *m*.
- Score transition using perceptron:

$$\underset{y \in \mathrm{GEN}(x)}{\operatorname{argmax}} \sum_{j=1}^{m} \mathbf{v}(y_j) \cdot \phi(x, y_1 \ldots y_{j-1}).$$

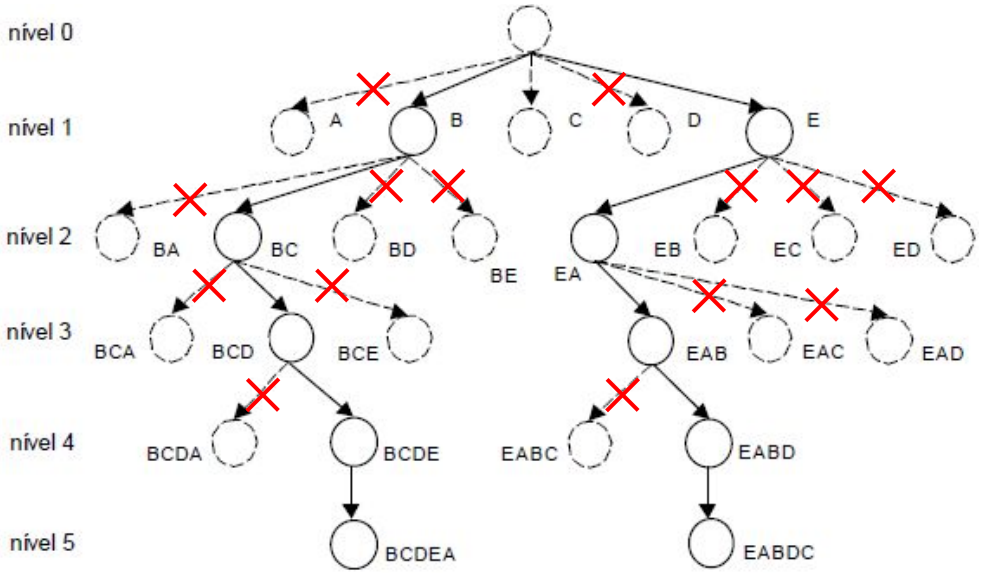


Figura 1 – Árvore de busca utilizando o *beam search*

3. Structured Perceptron Training + Beam Search

Problem: Greedy algorithms are unable to look beyond one step ahead, or recover from incorrect decisions.

Solution: Look forward -- search the tree of possible transition sequences.

- Keep track of *K* top partial transition sequences up to depth *m*.
- Score transition using perceptron:

$$\underset{y \in \mathrm{GEN}(x)}{\operatorname{argmax}} \sum_{j=1}^{m} \mathbf{v}(y_j) \cdot \phi(x, y_1 \ldots y_{j-1}).$$

Feature vector

Possible transition sequences

Perceptron parameter vector

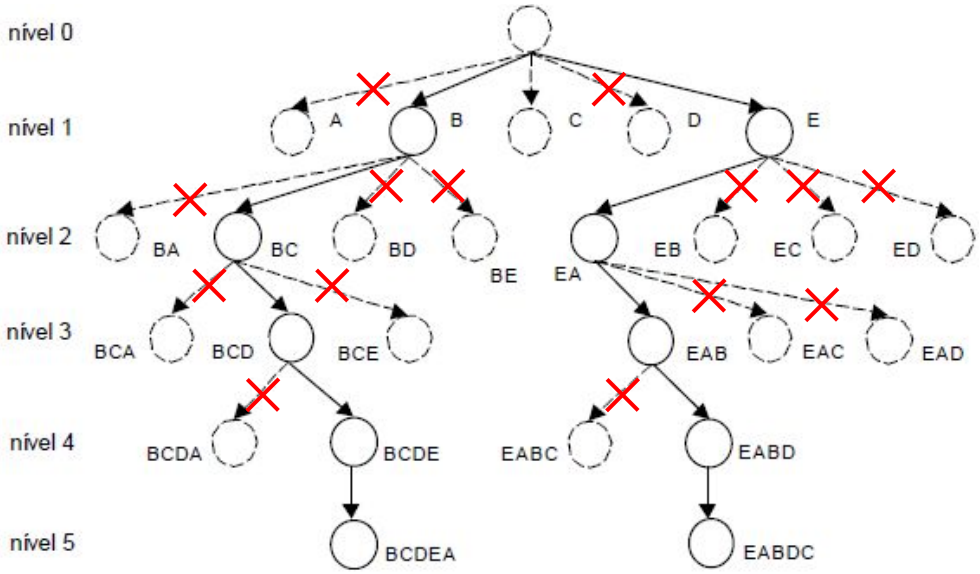


Figura 1 – Árvore de busca utilizando o *beam search*

Conclusions

Method	UAS	LAS (PTB WSJ SD 3.3
Chen & Manning 2014	92.0	89.7
Weiss et al. 2015	93.99	92.05
Andor et al. 2016	94.61	92.79

- Identify specific flaws in existing models (greedy algorithms) and solve them. In this case, with:
  - More data
  - Better tuning
  - Structured perceptron and beam search

- Final step to **SyntaxNet**: Andor et al. (2016) solve the “Label Bias Problem” using Global Normalization