

Reinforcement Learning Approach for Transition- Based Parsing

Yuqing Zhang and Kalki Seksaria

Reinforcement Learning (RL)

- RL with **global reward** allows long-term planning & frames parsing as a structured prediction problem
- RL samples a larger subset of possible configuration states

Two-Model Experimental Setup

- Parsing is a computationally heavy task
- Solution:
 1. Build **toy parser** (log-linear model w. ~1650 features) as a sandbox for ideas
 2. Extend successful toy parser ideas to **modify the neural parser** (Chen & Manning)

Data & Benchmarks

- Universal Dependencies for English
- Evaluation Metric: unlabeled arc score (**UAS**)
- Benchmarks:
 - Toy parser: **local vs. global** reward
 - Neural parser: **original vs. modified**

RL Framework

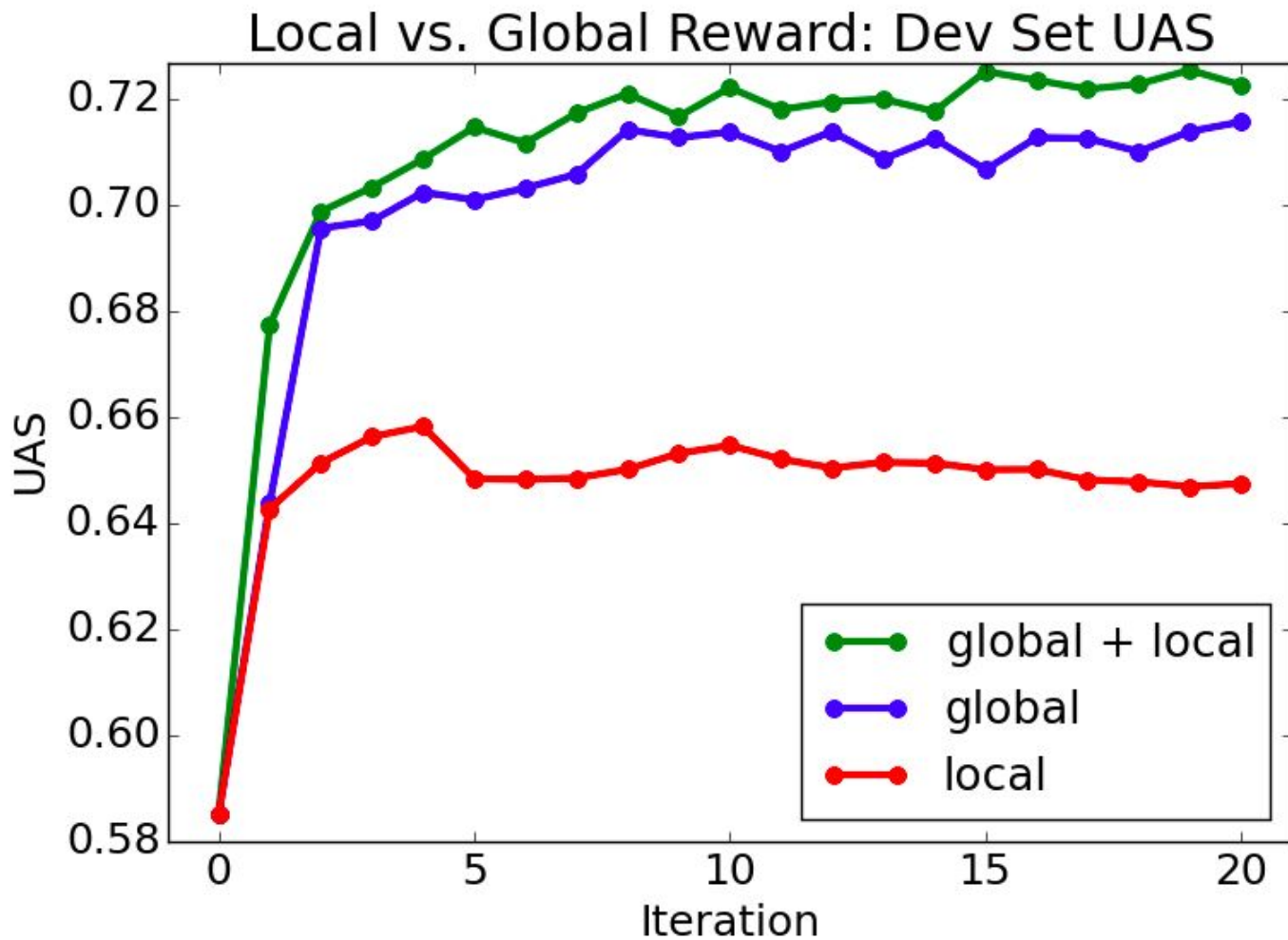
1. Initialize θ with “Oracle” training
2. **for** i in 1, 2, 3, ... NumIter **do**
 for Sentence S in TrainSet **do**
 $\text{config} \leftarrow \text{InitConfig}(S), \Delta_{\text{global}} \leftarrow 0$
 while NotDone(config) **do**
 $\text{trans} \sim \text{Policy}(\text{config}; \theta)$
 $r_{\text{local}} \leftarrow \text{LocalReward}(\text{trans}), \Delta_{\text{local}} \leftarrow \text{LocalDelta}(\text{config}, \text{trans})$
 $\theta \leftarrow \theta + r_{\text{local}} \times \Delta_{\text{local}}$ (toy-only)
 $\Delta_{\text{global}} \leftarrow \Delta_{\text{global}} + \Delta_{\text{local}}$
 $\text{config} \leftarrow \text{UpdateConfig}(\text{config}, \text{trans})$
 $r_{\text{global}} \leftarrow \text{GlobalReward}(\text{config})$
 $\theta \leftarrow \theta + r_{\text{global}} \times \Delta_{\text{global}}$

Toy Parser Reward Functions

$$\text{Local Reward} = \begin{cases} +1.0 & \text{if correct arc} \\ -0.5 & \text{if incorrect arc} \\ 0 & \text{o.w.} \end{cases}$$

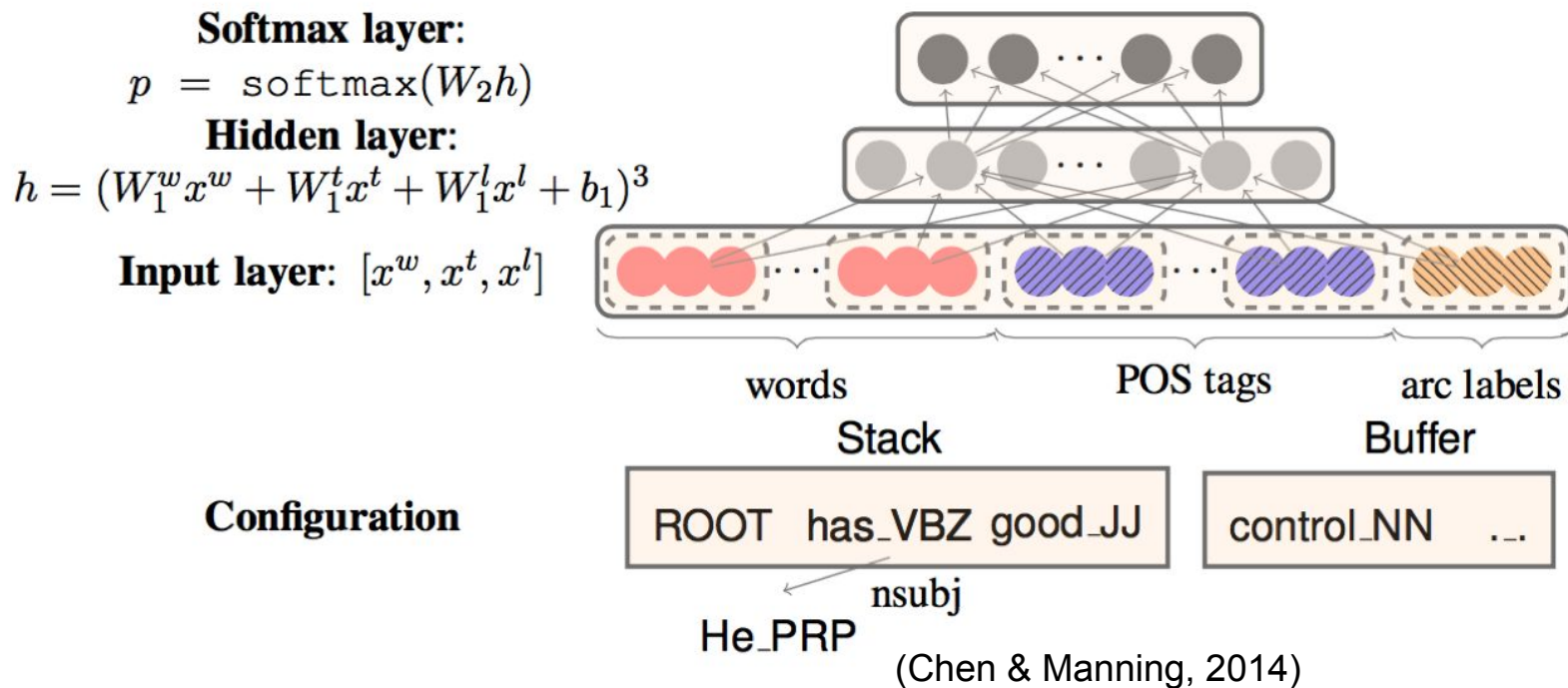
$$\text{Global Reward} = \begin{cases} \text{Accuracy} & \text{if Accuracy} > \text{Threshold} \\ 0 & \text{o.w.} \end{cases}$$

Toy Results

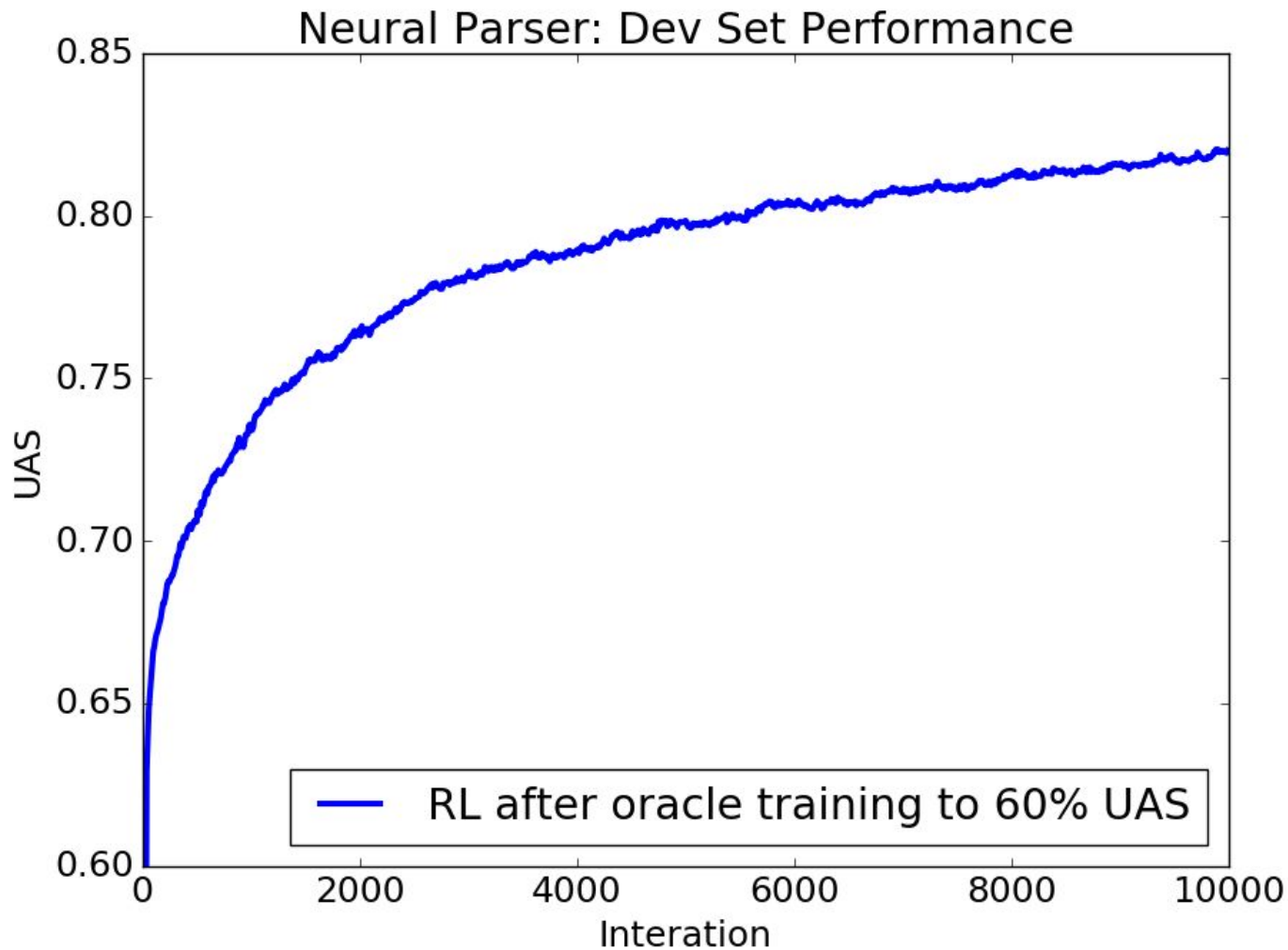


Neural Parser (Chen & Manning 2014)

- Per-arc (**local**) objective function
 - Treats each config state as separate problem



Neural Results



Test Set Results (UAS)

- Toy Parser:
 - Local Reward: 63.6%
 - Global + Local Reward: 72.1%
- Neural Parser:
 - Original: 83.4% (1000 oracle iter)
 - Modified with RL: 81.6% (10,000 RL iter)