

CS 224n Assignment #3: Dependency Parsing

Dongjin Na

1 Machine Learning & Neural Networks

- (a)
 - i. m(Momentum) value reflects the previous gradient so it accelerates the learning rate or not. The low variance may have benefit to avoid over-fitting issue.
 - ii. Parameters with small gradients of loss function will get larger updates because \sqrt{v} becomes small. It helps to turn accelerate convergence.
- (b)
 - i. γ is the same as $\frac{1}{1-p_{drop}}$.
 - ii. In evaluation, forward path should be deterministic.

2 Neural Transition-Based Dependency Parsing

- (a) Here are steps.

Stack	Buffer	New dependency	Transition
[ROOT]	[I, parsed, this, sentence, correctly]		Initial Configuration
[ROOT, I]	[parsed, this, sentence, correctly]		SHIFT
[ROOT, I, parsed]	[this, sentence, correctly]		SHIFT
[ROOT, parsed]	[this, sentence, correctly]	I \leftarrow parsed	LEFT-ARC
[ROOT, parsed, this]	[sentence, correctly]		SHIFT
[ROOT, parsed, this, sentence]	[correctly]		SHIFT
[ROOT, parsed, sentence]	[correctly]	this \leftarrow sentence	LEFT-ARC
[ROOT, parsed]	[correctly]	parsed \rightarrow sentence	RIGHT-ARC
[ROOT, parsed, correctly]	[]		SHIFT
[ROOT, parsed]	[]	parsed \rightarrow correctly	RIGHT-ARC
[ROOT]	[]	ROOT \rightarrow parsed	RIGHT-ARC

- (b) $2n$ steps because at least two steps are required to happen any transitions.
- (e) Achieved UAS of 88.80 on the dev set and UAS of 89.17 on the test set.