

cs224n Assignment #3: Dependency Parsing

1. Machine Learning & Neural Networks

(a) Adam Optimizer

- As using weighted average of existing gradient and newly calculated gradient, it can minimize the effect of potential outlier and prevent varying the update much.
This low variance helps optimizing the loss function faster.
- When $\mathbf{v} > 1$, adaptive learning rate makes the update slower, and vice versa.
Adaptive learning rate prevents overshooting from steep gradient and helps escaping the flat gradient faster.

(b) Dropout

- Because $\mathbf{h}_{\text{drop}} = \gamma \mathbf{d} \circ \mathbf{h} = \gamma \mathbb{1}_{1-p_{\text{drop}}} \circ \mathbf{h}$, $\mathbb{E}_{p_{\text{drop}}}[\mathbf{h}_{\text{drop}}]_i = \gamma(1 - p_{\text{drop}})h_i = h_i$
 $\therefore \gamma = \frac{1}{1 - p_{\text{drop}}}$
- As a part of regularization technique, we use dropout to keep some elements in weight matrixes zero so that it can prevent overfitting on the training data.
After finishing training, no more dropout is required to get the best result from the model as there is no overfitting on test data.

2. Neural Transition-Based Dependency Parsing

(a) The sequence of transitions is as follows.

No	Stack	Buffer	New dependency	Transition
0	[ROOT]	[I, parsed, this, sentence, correctly]		Initial Configuration
1	[ROOT, I]	[parsed, this, sentence, correctly]		SHIFT
2	[ROOT, I, parsed]	[this, sentence, correctly]		SHIFT
3	[ROOT, parsed]	[this, sentence, correctly]	parsed \rightarrow I	LEFT-ARC
4	[ROOT, parsed, this]	[sentence, correctly]		SHIFT
5	[ROOT, parsed, this, sentence]	[correctly]		SHIFT
6	[ROOT, parsed, sentence]	[correctly]	sentence \rightarrow this	LEFT-ARC
7	[ROOT, parsed]	[correctly]	parsed \rightarrow sentence	RIGHT-ARC
8	[ROOT, parsed, correctly]	\emptyset		SHIFT
9	[ROOT, parsed]	\emptyset	parsed \rightarrow correctly	RIGHT-ARC
10	[ROOT]	\emptyset	ROOT \rightarrow parsed	RIGHT-ARC

- (b) $2n$ steps in total to parse a sentence with n words, because SHIFT and ARC are needed to parse for each word.

```

a3 -- -zsh -- 80x26
1445850it [00:00, 48911050.66it/s]
- dev UAS: 88.61
New best dev UAS! Saving model.

Epoch 9 out of 10
100% | 1848/1848 [02:27<00:00, 12.54it/s]
Average Train Loss: 0.06919040547046827
Evaluating on dev set
1445850it [00:00, 47493749.86it/s]
- dev UAS: 88.49

Epoch 10 out of 10
100% | 1848/1848 [02:28<00:00, 12.47it/s]
Average Train Loss: 0.0661256366078762
Evaluating on dev set
1445850it [00:00, 50448256.27it/s]
- dev UAS: 88.46

=====
TESTING
=====
Restoring the best model weights found on the dev set
Final evaluation on test set
2919736it [00:00, 69603566.97it/s]
- test UAS: 89.05
Done!

```

- (e) dev UAS = 88.61, loss = 0.0661
test UAS = 89.05

- (f)
 - i.
 - Error type: Verb Phrase Attachment Error
 - Incorrect dependency: wedding \rightarrow fearing
 - Correct dependency: heading \rightarrow fearing
 - ii.
 - Error type: Coordination Attachment Error
 - Incorrect dependency: makes \rightarrow rescue
 - Correct dependency: rush \rightarrow rescue
 - iii.
 - Error type: Prepositional Phrase Attachment Error
 - Incorrect dependency: named \rightarrow Midland
 - Correct dependency: guy \rightarrow Midland
 - iv.
 - Error type: Modifier Attachment Error
 - Incorrect dependency: elements \rightarrow most
 - Correct dependency: crucial \rightarrow most