CS224n HW2

Xinyu Tan

August 22, 2017

1 Neural Transition-based Dependency Parsing

(a)

stack	buffer	new dependency	transition
[root]	[I, parsed, this, sentence, correctly]		Initial Configuratio
[root, I]	[parsed, this, sentence, correctly]		SHIFT
[root, I, parsed]	[this, sentence, correctly]		SHIFT
[root, parsed]	[this, sentence, correctly]	parsed $\rightarrow$ I	LEFT-ARC
[root, parsed, this]	[sentence, correctly]		SHIFT
[root, parsed, this, sentence]	[correctly]		SHIFT
[root, parsed, sentence]	[correctly]	sentence $\rightarrow$ this	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)

The sentence will be parsed in $2n$ times. Each word will be pushed into stack once, and each word only depends on one other word. Therefore, the process is in $O(n)$ time complexity.

(f)

We need to satisfy: $\mathbb{E}_{p_{\text{drop}}}[\boldsymbol{h}_{\text{drop}}]_i = \gamma(1 - p_{\text{drop}})\boldsymbol{h}_i = \boldsymbol{h}_i$, then we have:

$$\gamma = \frac{1}{1 - p_{\text{drop}}}$$

(g)

(i)

(ii)

2 Recurrant neural networks: Language Modeling

(a)

Perplexity:

$$PP^{(t)}\left(y^{(t)}, \hat{y}^{(t)}\right) = \frac{1}{y_k^{(t)} \hat{y}_k^{(t)}} = \frac{1}{\hat{y}_k^{(t)}}$$

Cross-entropy loss:

$$J^{(t)}(\theta) = -y_k^{(t)} \log \hat{y}_k^{(t)} = -\log \hat{y}_k^{(t)}$$

Then, it is easy to derive that

$$PP^{(t)}\left(y^{(t)}, \hat{y}^{(t)}\right) = e^{J^{(t)}(\theta)}$$

Therefore, minimizing perplexity equals to minimizing the cross-entropy.

For a vocabulary of $|V| = 10000$ words, if the model is completely random, then the perplexity will be 10000, and then the cross entropy will be $\log 10000 = 9.21$.

(b)

The derivatives:

$$\frac{\partial J^{(t)}}{\partial \boldsymbol{b}_2} = \frac{\partial J^{(t)}}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{b}_2} = \hat{\boldsymbol{y}}^{(t)} - \boldsymbol{y}^{(t)}$$