به نام خدا

درس مبانی پردازش زبان و گفتار

تمرین سری دوم

مدرس درس:

جناب آقای دکتر اعتمادی

تهیه شده توسط:

الناز رضایی ۹۸۴۱۱۳۸۷

تاریخ ارسال: ۱۴۰۱/۱۲/۲۶

پاسخ (a)

i . In optimization algorithms like gradient descent, the momentum term (m) is used to reduce the variance in updates made to the model's parameters between iterations. This low variance helps the model avoid overshooting the minimum or getting stuck in local minima. By maintaining a certain level of momentum in the direction of the gradient, the model can converge faster and more smoothly towards the optimal solution.

ii . Adam optimizer divides the update by the square root of the exponential moving average of the squared gradients (i.e., v) for each parameter in the model. This normalization factor helps to scale down the updates for parameters that have large gradients and scale up the updates for parameters that have small gradients.

The parameters with smaller v values will receive larger updates because the normalization factor will be smaller for those parameters. Conversely, parameters with larger v values will receive smaller updates because the normalization factor will be larger for those parameters.

This approach helps with learning because it allows the optimizer to adapt the learning rate to each parameter based on the gradient's magnitude. Parameters that have large gradients will be updated more slowly, whereas parameters that have small gradients will be updated more quickly. This helps to prevent the optimizer from overshooting the minimum and converging faster to the optimal solution.

i . γ represents the probability of keeping a unit during training, whereas p_{drop} represents the probability of dropping out a unit during training. Therefore, the probability of keeping a unit is simply the complement of the probability of dropping out a unit, which is 1 minus p_{drop}. So we have:

$$\gamma = \frac{1}{1 - pdrop}$$

ii . During training, dropout should be applied because it can prevent overfitting by regularizing the neural network. Overfitting occurs when the network learns to fit the training data too well, including the noise in the data, which leads to poor performance on new and unseen data. Dropout can help prevent this by introducing noise and making the network more robust to small variations in the input.

However, dropout should not be applied during evaluation or testing because it can result in inconsistent and unstable predictions. During evaluation, the goal is to get the most accurate prediction possible, and dropping out neurons randomly can interfere with this goal. Instead, the full network should be used for evaluation to ensure that all the learned features are utilized in making the final prediction.

پاسخ (a)

Stack	Buffer	New dependency	Transition
[ROOT]	[I, attended, lectures, in, the, NLP, class]		Initial Configuration
[ROOT, I]	[attended, lectures, in, the, NLP, class]		SHIFT
[ROOT, I, attended]	[lectures, in, the, NLP, class]		SHIFT
[ROOT, attended]	[lectures, in, the, NLP, class]	attended $\rightarrow$ I	$LEFT-ARC$
[ROOT, attended, lectures]	[in, the, NLP, class]		SHIFT
[ROOT, attended]	[in, the, NLP, class]	attended $\rightarrow$ lectures	$Right-ARC$
[ROOT, attended, in]	[the, NLP, class]		SHIFT
[ROOT, attended, in, the]	[NLP, class]		SHIFT
[ROOT, attended, in, the, NLP]	[class]		SHIFT
[ROOT, attended, in, the, NLP, class]	[]		SHIFT
[ROOT, attended, in, the, class]	[]	class $\rightarrow$ NLP	$LEFT-ARC$
[ROOT, attended, in, class]	[]	class $\rightarrow$ the	$LEFT-ARC$
[ROOT, attended, class]	[]	class $\rightarrow$ in	$LEFT-ARC$
[ROOT, attended]	[]	attend $\rightarrow$ class	$Right-ARC$
[ROOT]	[]	Root $\rightarrow$ attend	$Right-ARC$

پاسخ (b)

2n times. Because we move all words from buffer to stack, which takes n times, and then, remove all of them from stack (again n times).

پاسخ (e)

```
Epoch 10 out of 10
100%|
Average Train Loss: 0.029609152744322596
Evaluating on dev set
1445850it [00:00, 43808898.83it/s]
- dev UAS: 88.43


================================================================================
TESTING
================================================================================
Restoring the best model weights found on the dev set
Final evaluation on test set
2919736it [00:00, 66359569.45it/s]
- test UAS: 89.13
Done!
```

i . The university blocked the acquisition , citing concerns about the risks involved.

Error type: Verb Phrase Attachment Error

Incorrect dependency: acquisition → citing

Correct dependency: blocked → citing

ii . Many managers and traders had already left their offices early Friday afternoon.

Error type: Modifier Attachment Error

Incorrect dependency: left → early

Correct dependency: afternoon → early

iii . Investment Canada declined to comment on the reasons for the goverment decision.

Error type: Prepositional Phrase Attachment Error

Incorrect dependency: declined → decision

Correct dependency: reasons → decision

iv . People benefit from a separate move that affects three US car plants and one in Quebec.

Error type: Coordination Attachment Error

Incorrect dependency: affects → one

Correct dependency: plants → one

The benefit of using POS tags as features in a parser is that it can help disambiguate between the different possible parse trees for a sentence.

Furthermore, POS tags can also help to identify the grammatical relationship between words in a sentence, such as subject-verb-object relationships. This information can be used to construct a parse tree that accurately represents the underlying syntactic structure of the sentence.

Overall, using POS tags as features in a parser can improve its accuracy and ability to correctly parse sentences, which is essential for natural language processing applications such as machine translation, text-to-speech synthesis, and sentiment analysis.