

CreateMoMo

2017-10-08

CRF Layer on the Top of BiLSTM - 3

2.3 CRF loss function

The CRF loss function is consist of the real path score and the total score of all the possible paths. The real path should have the highest score among those of all the possible paths.

For example, if we have these labels in our dataset as shown in the table:

Label	Index
B-Person	0
I-Person	1
B-Organization	2
I-Organization	3
O	4
START	5
END	6

We also have a sentence which has 5 words. The possible paths could be:

- 1) START B-Person B-Person B-Person B-Person B-Person END
- 2) START B-Person I-Person B-Person B-Person B-Person END

- ...
- **10) START B-Person I-Person O B-Organization O END**
- ...
- N) O O O O O O O

Suppose every possible path has a score $P_{\{i\}}$ and there are totally N possible paths, the total score of all the paths is $P_{\{total\}} = P_1 + P_2 + \dots + P_N = e^{\{S_1\}} + e^{\{S_2\}} + \dots + e^{\{S_N\}}$, e is the mathematical constant e . (In section 2.4, we will explain how to calculate S_i , you can also treat it as a score of this path.)

If we say the 10th path is the real path, in other words, the 10th path is the gold standard labels provided by our training dataset. The score $P_{\{10\}}$ should be the one with the largest percentage in all the possible paths.

Given the equation which is also **the loss function** below, **during the training process, the parameter values of our BiLSTM-CRF model will be updated again and again to keep increasing the percentage of the score of the real path.**

$$\text{\$ Loss Function} = \frac{P_{\{\text{RealPath}\}}}{P_1 + P_2 + \dots + P_N}$$

Now, the questions are:

- 1) How to define the score of a path?
- 2) How to calculate the total score of all possible paths?
- 3) When we calculate the total score, do we have to list all the possible paths? (The answer for this question is NO.)

In the following sections, we will see how to solve these questions.

Next

2.4 Real path score

How to calculate the score of the true labels of a sentence.

2.5 The score of all the possible paths

How to calculate the total score of all the possible paths of a sentence with a step-by-step toy example.

(Sorry for my late update, I will try my best to squeeze time for updating the following sections.)

References

- [1] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K. and Dyer, C., 2016. Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360.
<https://arxiv.org/abs/1603.01360>

When you reprint or distribute this article, please include the original link address.

 Comments  Share

OLDER

CRF Layer on the Top of BiLSTM - 2

ARCHIVES

[October 2017](#)
[September 2017](#)

RECENT POSTS

[CRF Layer on the Top of BiLSTM - 3](#)
[CRF Layer on the Top of BiLSTM - 2](#)
[CRF Layer on the Top of BiLSTM - 1](#)