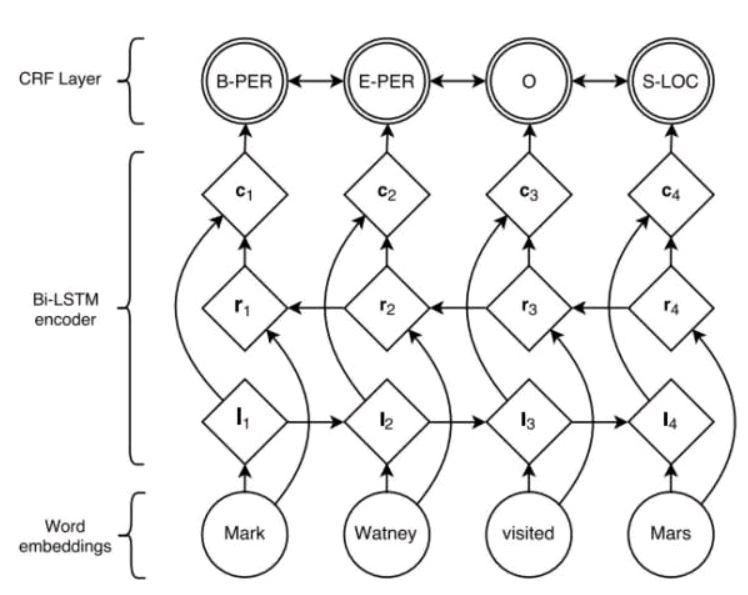
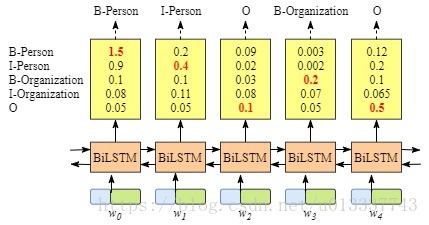
BiLSTM+CRF

整体架构图：



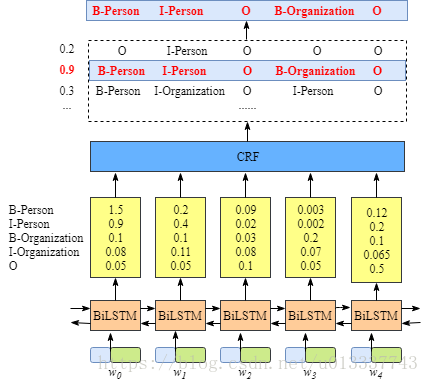
BiLSTM+Softmax：



双向lstm后接一个softmax层，输出各个label的概率。那为何还要加一个crf层呢？

我的理解是softmax层的输出是相互独立的，即虽然BiLSTM学习到了上下文的信息，但是输出相互之间并没有影响，它只是在每一步挑选一个最大概率值的label输出。这样就会导致如B-person后再接一个B-person的问题。而crf中有转移特征，即它会考虑输出label之间的顺序性，所以考虑用crf去做BiLSTM的输出层。

BiLSTM+CRF：



解释：

BiLSTM层的输出为每一个标签的预测分值，例如，对于单元w0,BiLSTM层输出的是1.5 (B-Person), 0.9 (I-Person), 0.1 (B-Organization), 0.08 (I-Organization) and 0.05 (O). 这些分值将作为CRF的输入。即使没有CRF层，我们也可以训练一个BiLSTM命名实体识别模型，由于BiLSTM的输出为单元的每一个标签分值，我们可以挑选分值最高的一个作为该单元的标签。例如，对于单元w0,“B-Person”有最高分值—— 1.5，因此我们可以挑选“B-Person”作为w0的预测标签。同理，我们可以得到w1——“I-Person”，w2—— “O” ，w3——“B-Organization”，w4——“O”。

        虽然我们可以得到句子x中每个单元的正确标签，但是我们不能保证标签每次都是预测正确的。例如，标签序列是“I-Organization I-Person” and “B-Organization I-Person”，很显然这是错误的。

CRF层可以为最后预测的标签添加一些约束来保证预测的标签是合法的。在训练数据训练过程中，这些约束可以通过CRF层自动学习到。

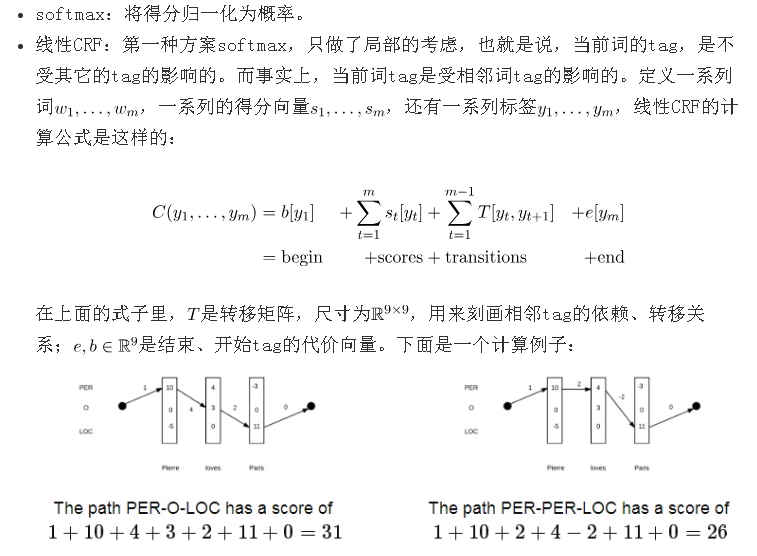
这些约束可以是：

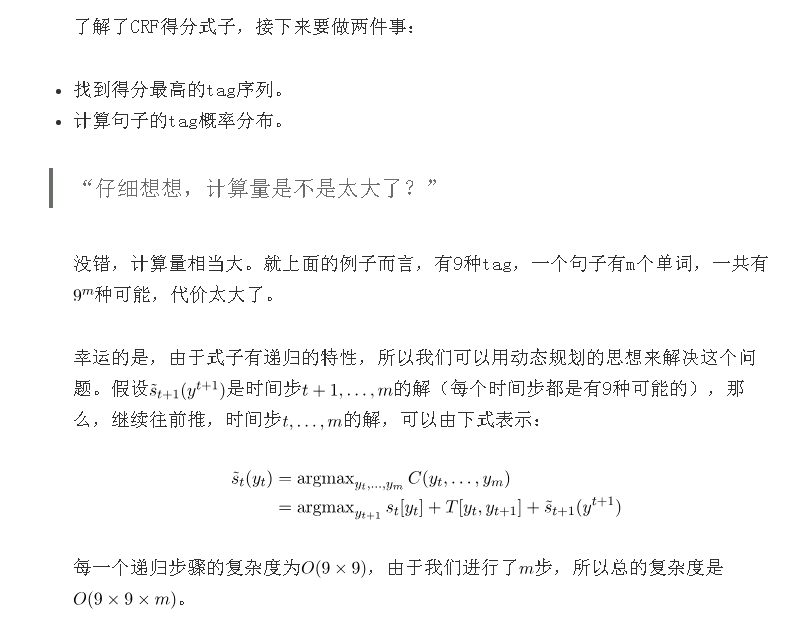
I：句子中第一个词总是以标签“B-“ 或 “O”开始，而不是“I-”

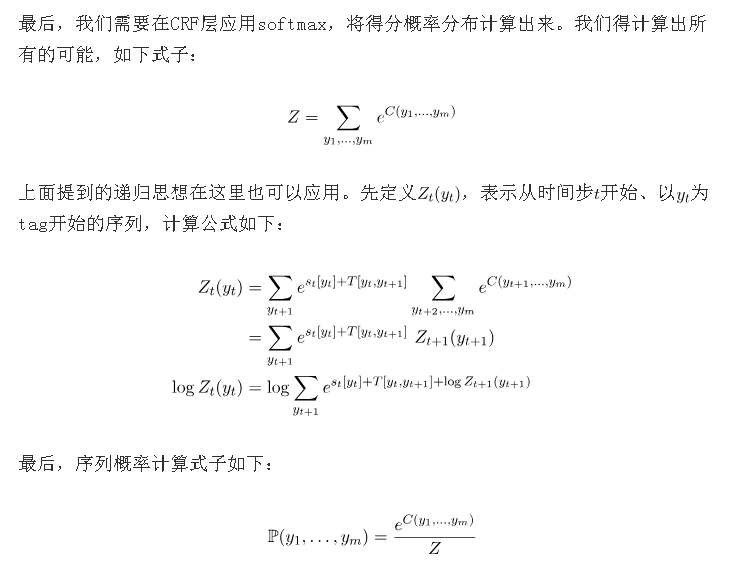
II：标签“B-label1 I-label2 I-label3 I-…”,label1, label2, label3应该属于同一类实体。例如，“B-Person I-Person” 是合法的序列, 但是“B-Person I-Organization” 是非法标签序列.

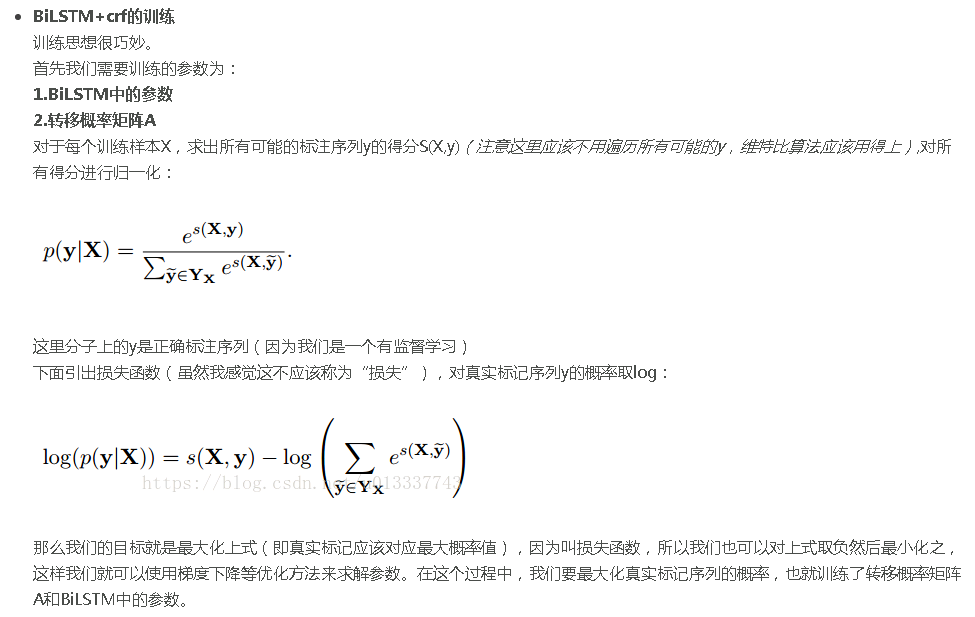
III：标签序列“O I-label” is 非法的.实体标签的首个标签应该是 “B-“ ，而非 “I-“, 换句话说,有效的标签序列应该是“O B-label”。

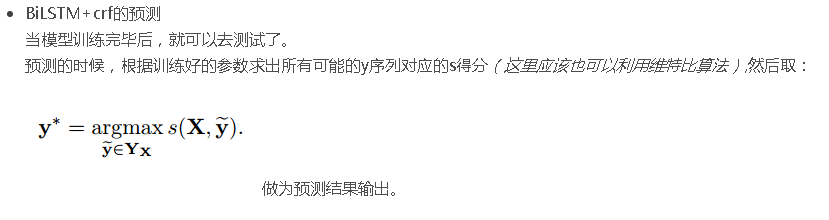
有了这些约束，标签序列预测中非法序列出现的概率将会大大降低。











参考资料：

<https://blog.csdn.net/bobobe/article/details/80489303>

<https://www.lookfor404.com/%E7%94%A8%E5%8F%8C%E5%90%91lstmcrf%E5%81%9A%E5%91%BD%E5%90%8D%E5%AE%9E%E4%BD%93%E8%AF%86%E5%88%AB%E9%99%84tensorflow%E4%BB%A3%E7%A0%81-ner%E7%B3%BB%E5%88%97%EF%BC%88%E5%9B%9B/>