Bi-LSTM-CRF模型大概分为三部分：输入层、隐藏层及输出层，如下图2所示，

图2 Bi-LSTM-CRF模型架构

1. **输入层**

大规模语料（中文）经过数据预处理、分词后，还需要通过Word Embedding（词嵌入）将词映射为向量。因为计算机无法对一个中文单词进行计算，将中文单词映射到向量空间，每个中文单词对应一个高维向量。向量之间的距离代表着两个单词语义的相似度，即语义相近的词对应的向量，在向量空间中的距离也是相近的。常用的词嵌入工具是谷歌的word2vec开源软件，它应用了CBoW模型和Skip-gram模型[3]，将分好词的大规模语料送入word2vec，经过训练调参，可以得到每个词对应的高维向量，接着，我们便可以使用这些具有语义信息的向量进行后续的计算了。这样做词嵌入既可以使不可计算的中文单词变得可计算，又在很大程度上保留了中文单词的语义信息。当然，不使用预训练好的词向量也可以，词向量会在训练初始的时候在一定数值范围内随机初始化，然后词向量会随着训练的过程不断更新，最终也能达到比较理想的效果。

除了单词本身的语义向量，其他的特征也可以以向量的形式一起嵌入模型，如词性标注（Part-Of-Speech）。每个中文单词都有对应的词性，如动词、名词、形容词等。在分词的步骤中，jieba等开源工具可以同时用隐马尔科夫模型（HMM）为每个分好的中文单词打上词性标注的标签。类似词嵌入，每一种词性可以映射为一个高维向量，然后与词向量拼接到一起送入模型中训练。在训练的过程中，这些向量会不断调整其数值，以拟合适应模型，提高模型性能。

在图2中，输入层负责接收词嵌入之后的高维向量。“有”、“慢性”、“咽炎”、“症状”这4个中文单词分别对应4个高维向量，输入到模型当中进行后续计算。

1. **隐藏层**

Bi-LSTM-CRF模型的隐藏层应用的是双向LSTM神经网络的神经元，这是一个时间序列模型，在句子中处于不同位置的中文单词可视为不同时刻的输入。隐藏层每个神经元结构如下图3所示：

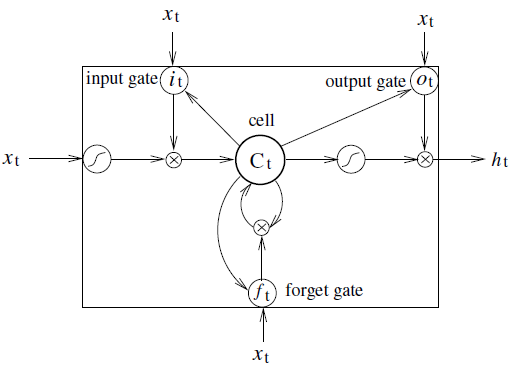


图3 LSTM神经元结构

如图所示，每一个时刻的LSTM神经元中都设置了输入门（input gate）、输出门（output gate）、遗忘门（forget gate）以及细胞状态（cell）。这些门主要负责接收输入层传过来的词向量、当前细胞状态量以及上一时刻神经元的输出量，在训练的过程中，门的参数会不断进行调整，来决定哪些信息需要继续往下一时刻传递、传递多少，哪些信息需要被遗忘、忘掉多少，由此来控制有用的信息流动。这个过程就是模型自动学习出隐藏的特征的过程，这些特征同样以高维向量的形式存在，作为输出层的输入，为解码出每个中文单词的标签做准备。

由于传统LSTM神经网络只能控制信息的前向流入程度，即在句子中当前位置的单词只能获取在其之前的单词的信息，但实际上该词位置之后的单词也同样非常重要，所以本课题还应用了双向LSTM网络[4]。如图2所示，隐藏层中上面一层用了前向LSTM，下面一层用了后向LSTM，将这2个相互独立的LSTM网络输出的隐向量拼接在一起，便构成了输出层最终的输入。

1. **输出层**

输出层接收隐藏层计算出来的隐向量，通过一个条件随机场（CRF）层，可以计算出这句话的所有标签组合的概率集合，然后我们只需要挑取概率最大的标签序列，即可作为最终预测出来的标签序列。如图2中输出层所示，将每个时刻的隐向量输入到CRF层，CRF模型经过综合计算整个句子不同位置的词向量与状态转移，得到所有可能的标签序列的概率。挑选出最大可能的标签序列“O”、“B-DIS”、“E-DIS”及“O”分别对应输入的4个中文单词“有”、“慢性”、“咽炎”、“症状”，再根据标注规则“B-DIS”代表“Disease”类实体的开头（Begin），“E-DIS”代表“Disease”类实体的结尾（End），“O”则代表不是实体（Other），最终预测抽取出的实体即为“慢性咽炎”，属于“Disease”类。

[1] Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. In *Computer Science*, *2015.*

[2] J. Lafferty, A. McCallum, and F. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of ICML, 2001*.

[3] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality*.* In *Proc. NIPS, 2013*.

[4] A. Graves, A. Mohamed, and G. Hinton. 2013. Speech Recognition with Deep Recurrent Neural Networks. In *ICASSP, 2013*.