1. 数据预处理：对样本进行分割。
2. 预训练字向量：中文CNKI的医疗摘要和CCKS提供的中文电子医疗记录信息（总共有156w条文本），训练词向量。
3. 训练笔画向量：模型需要研究，数据来源未知。具体来讲，首先在语料词汇表中每个字的笔画序列通过handian网站获得。然后，笔画的embedding向量被输入到卷积层。使用max pooling层来抽取卷积层中的特征。接着，这些特征向量被输入到基于BILSTM的双向语言模型中。在使用上述提及的无标注数据训练双向语言模型后，笔画elmo抽 取biLM的中间层表示，然后接上全连接获取512维的语境字embedding。
4. Bilstm-crf模型：输入为句子的字向量拼接上笔画的elmo向量。
5. 迁移学习：原数据ccks2018的数据集，目标数据ccks2019数据集。模型结构：BILSTM-CRF。使用了3种结构的模型训练。

总结：主体模型使用的BILSTM-CRF，在该基础上输入特征增加了字的笔画向量作为输入，使用迁移学习充分训练。预处理样本：分割句子长度，后处理部分去除空格，匹配括号。