1、环境的安装: (本文只说明conda方式的安装，非conda环境基本相同)

初始化python环境：

#conda create -n ner python==3.6 注：python为3.6

#conda activate ner 注：激活ner环境

安装tensorflow环境：

#pip install tensorflow==1.12 注:tensorflow为1.12版本，其它版本可能需要验证国；

2、源码的下载 BERT-BiLSTM-CRF-NER

#git clone <https://github.com/macanv/BERT-BiLSTM-CRF-NER>

#cd BERT-BiLSTM-CRF-NER/

#python3 setup.py install

3、谷歌BERT模型

链接： https://github.com/google-research/bert

链接: https://pan.baidu.com/s/14PxSs-ntIFkqAllS5GM8wA 密码: sfwd

下载后解压到BERT-BiLSTM-CRF-NER/目录

解压后会建立chinese\_L-12\_H-768\_A-12-3目录，目录下会有5个文件：

vocab.txt，bert\_config.json，bert\_model.ckpt.index， bert\_model.ckpt.data-00000-of-00001，bert\_model.ckpt.meta；

vocab.txt是模型的词典，这个文件会经常要用到，后面会讲到。

bert\_config.json是BERT的配置(超参数)，比如网络的层数，通常我们不需要修改，但是也会经常用到。

bert\_model.ckpt\*，这是预训练好的模型的checkpoint

4、训练文本准备：

建立训练数据目录

#mkdir data

目录中需要有三个文件：train.txt, dev.txt, test.txt

这三个文件为训练数据，验证数据和测试数据

该部分内容可以使用人民网的数据

训练数据的标准格式为BIO。

BERT-BiLSTM-CRF-NER中只做了三个实体的标识，ORG、PER、LOC即组织、人员和地点，如果需要训练自已的实体，可以在程序中增加:

#vi /root/BERT-BiLSTM-CRF-NER/bert\_base/train/bert\_lstm\_ner.py

修改703行，增加实体标识名称；

5、训练模型

bert-base-ner-train \

-data\_dir ./data \

-output\_dir ./output \

-init\_checkpoint ./chinese\_L-12\_H-768\_A-12/bert\_model.ckpt \

-bert\_config\_file ./chinese\_L-12\_H-768\_A-12/bert\_config.json \

-vocab\_file ./chinese\_L-12\_H-768\_A-12/vocab.txt -batch\_size 10

data\_dir：训练数据存放目标，其中目录中有三个文件：train.txt dev.txt test.txt,分别是训练数据、验证数据和测试数据；

out\_dir：输出模型的地址

init\_checkpoint：其为google提供的bert预训练模型

bert\_config\_file：为google提供的bert模型的配置文件

vocab\_file：为google提供的bert模型的配置文件

batch\_size：其为运行的批次空间的大小，其根据GPU的内存，如果显示OOM错误，需要调小这个数值

5、使用模型

关于汽车数据的标记：

数据分类：

BRAND：品牌或公司 如:大众 奔驰等

VEHICLE：汽车名称如：迈腾、雅阁等

POWER：动力，如1.4T, 2.0T

GARE：汽车的装备，如：四驱、自动档等

MODLE：汽车的型号：如：豪华型、尊贵型等

SPARE：汽车的配件：如：激光大灯，LED

FOCUS：客人的关注点，如：置换分期、二手车

PRICE：谈论的价格，如：金额及保险等

PAY：支付的方式，如：微信、转帐、支付宝等

PHONE：电话号码