# 环境

linux+Anaconda3+python3.7+CUDA8.0+cudnn6.0

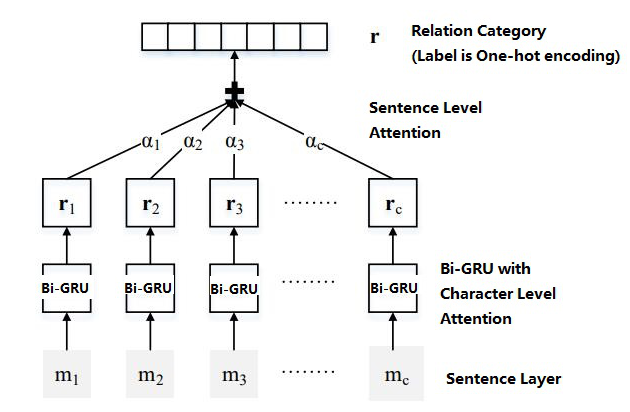
安装指导: <https://blog.csdn.net/qq_38901147/article/details/90049666>

# 数据集

一共分为5类标签，测试集1424条，训练集6174条

详见附件

# 模型



其中，此模型包含四个部分，第一层是句子层也叫作词嵌入层，目的是获取词向量以及位置信息。第二层为BiGRU与词注意力机制结合层，BiGRU层是提取句子的全局信息，而添加了字的注意力机制可以获取与实体关系比较大字的特征。第三层为句子注意力层，句子注意力层的作用是能够更好的提取关系这个类的特征。第四层为softmax分类器进行类别划分。

# 操作步骤

## 调试参数

**第一步：导入数据集和标签文件**

进入origin\_data文件夹下，导入relation2id.txt，test.txt，train.txt

注意：如果更改数据集，关系，必须严格按照格式进行标注和关系更改

**第二步：进入虚拟环境**

root@gpu:/sk/relation\_extract/Relationship-extraction-master# source activate

(base) root@gpu:/sk/relation\_extract/Relationship-extraction-master# conda env list

# conda environments:

#

base \* /usr/local/anaconda3

tensorflow\_py37 /usr/local/anaconda3/envs/tensorflow\_py37

tensorflow\_py37\_gpu /usr/local/anaconda3/envs/tensorflow\_py37\_gpu

tensorflow\_py37\_tf14 /usr/local/anaconda3/envs/tensorflow\_py37\_tf14

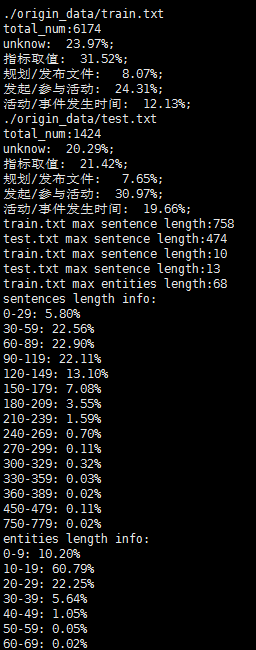
(base) root@gpu:/sk/relation\_extract/Relationship-extraction-master# conda activate tensorflow\_py37\_gpu

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master#

**第三步：数据预处理**

1. 查看数据集基本信息，包含了测试集，训练集数目，各个指标所占的比例，固定句子长度所占的比例，实体对长度所占的比例

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u dataset\_info.py > logs/dataset.log 2>&1 &



2、生成格式化的测试集与训练集

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u initial.py > logs/initial.log 2>&1 &

[1] 19988

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master#

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master#

查看日志：

(tensorflow\_py37\_gpu)root@gpu:/sk/relation\_extract/Relationship-extraction-master#cat logs/initial.log

**第四步：执行模型**

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py > logs/train.log 2>&1 &

[1] 21040

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master#

**第五步：查看模型参数文件**

进入model目录下，查看对应的文件

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master/model# ll

total 734480

drwxr-xr-x 2 root root 4096 6月 9 14:47 ./

drwxr-xr-x 11 root root 4096 6月 9 13:36 ../

-rw-r--r-- 1 root root 25043676 6月 9 14:44 ATT\_GRU\_model-8000.data-00000-of-00001

-rw-r--r-- 1 root root 1901 6月 9 14:44 ATT\_GRU\_model-8000.index

-rw-r--r-- 1 root root 14530283 6月 9 14:44 ATT\_GRU\_model-8000.meta

-rw-r--r-- 1 root root 25043676 6月 9 14:45 ATT\_GRU\_model-8100.data-00000-of-00001

-rw-r--r-- 1 root root 1901 6月 9 14:45 ATT\_GRU\_model-8100.index

-rw-r--r-- 1 root root 14530283 6月 9 14:45 ATT\_GRU\_model-8100.meta

-rw-r--r-- 1 root root 25043676 6月 9 14:45 ATT\_GRU\_model-8200.data-00000-of-00001

-rw-r--r-- 1 root root 1901 6月 9 14:45 ATT\_GRU\_model-8200.index

-rw-r--r-- 1 root root 14530283 6月 9 14:45 ATT\_GRU\_model-8200.meta

-rw-r--r-- 1 root root 25043676 6月 9 14:45 ATT\_GRU\_model-8300.data-00000-of-00001

## 获取最优模型参数

**第一步：执行测试集**

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u ，model\_test.py > logs/test.log 2>&1 &

查看模型ID文件

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# cd data/

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master/data# ll

total 313132

drwxr-xr-x 2 root root 4096 5月 27 14:19 ./

drwxr-xr-x 11 root root 4096 6月 9 13:36 ../

-rw-r--r-- 1 root root 8928 6月 9 14:32 allans.npy

-rw-r--r-- 1 root root 63172 6月 9 14:32 metadata.tsv

-rw-r--r-- 1 root root 4 6月 9 14:59 model\_id.txt

-rw-r--r-- 1 root root 56128 6月 9 14:32 testall\_pos1.npy

-rw-r--r-- 1 root root 56128 6月 9 14:32 testall\_pos2.npy

-rw-r--r-- 1 root root 56128 6月 9 14:32 testall\_word.npy

-rw-r--r-- 1 root root 168128 6月 9 14:32 testall\_x.npy

-rw-r--r-- 1 root root 9728 6月 9 14:32 testall\_y.npy

-rw-r--r-- 1 root root 2423 6月 9 14:32 test\_q&a.txt

-rw-r--r-- 1 root root 29973207 6月 9 14:32 train\_pos1.npy

-rw-r--r-- 1 root root 29973207 6月 9 14:32 train\_pos2.npy

-rw-r--r-- 1 root root 1282330 6月 9 14:32 train\_q&a.txt

-rw-r--r-- 1 root root 39896804 6月 9 14:32 train\_word.npy

-rw-r--r-- 1 root root 207896804 6月 9 14:32 train\_x.npy

-rw-r--r-- 1 root root 4712288 6月 9 14:32 train\_y.npy

-rw-r--r-- 1 root root 6446928 6月 9 14:32 vec.npy

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master/data#

## 关系抽取

**第一步：待抽取的文件放置data路径下，并更名为pre\_text**

注意:必须严格按照格式进行

**第二步：执行model\_predict.py**

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_predict.py > logs/predict.log 2>&1 &

查看日志：

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# cat logs/predict.log

**第三步：获取结果文件result.csv**

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# cd out

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master/out# ll

total 164

drwxr-xr-x 2 root root 4096 6月 9 15:50 ./

drwxr-xr-x 11 root root 4096 6月 9 13:36 ../

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8000.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8100.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8200.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8300.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8400.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8500.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8600.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8700.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8800.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_8900.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9000.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9100.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9200.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9300.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9400.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9500.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9600.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9700.npy

-rw-r--r-- 1 root root 4528 6月 9 14:59 allprob\_iter\_9800.npy

-rw-r--r-- 1 root root 31 6月 9 15:50 result.csv

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master/out#

# 命令总结

1. 调整训练集的训练次数

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py -num\_train\_epochs=20 > logs/train.log 2>&1 &

1. 调整Batchsize

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py -batch\_size=10 > logs/train.log 2>&1 &

1. 调整正则率

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u GRU\_train.py -regularizer=0.01 > logs/train.log 2>&1 &

注意：正则率一般非常的小，默认是0.001

1. 调整开始保存模型的step点

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u GRU\_train.py -begin\_save\_steps=8000> logs/train.log 2>&1 &

1. 调整模型为LSTM模型

(tensorflow\_py37\_gpu\_temp) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py -cell='lstm' > logs/train.log 2>&1 &

(tensorflow\_py37\_gpu\_temp) root@gpu:/sk/relation\_extract/Relationship-extraction-master/data# nohup python -u model\_test.py -cell='lstm' > logs/test.log 2>&1 &

(tensorflow\_py37\_gpu\_temp) root@gpu:/sk/relation\_extract/Relationship-extraction-master#nohup python -u model\_predict.py -cell='lstm' > logs/predict.log 2>&1 &

1. 关系的单向性作为影响因子加入到模型中

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py -weight='relation' > logs/train.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_test.py -weight='relation' > logs/test.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_predict.py -weight='relation' > logs/ predict.log 2>&1 &

1. Sentence represent的计算方式（max和sum两种方式）

Max：

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py -repre='max' > logs/train.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_test.py -repre='max' > logs/test.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_predict.py -repre='max' > logs/ predict.log 2>&1 &

Sum:

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py -repre='vector' > logs/train.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_test.py -repre='vector' > logs/test.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_predict.py -repre='vector' > logs/ predict.log 2>&1 &

1. 支持的输入语句的最大长度

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u initial.py -max\_sentence\_len=200 > logs/initial.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py -max\_sentence\_len=200 > logs/train.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_test.py -max\_sentence\_len=200 > logs/test.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_predict.py -max\_sentence\_len=200> logs/ predict.log 2>&1 &

注意：max\_sentence\_len的长度设定可以参考4.1章节，第一步中数据集的基本信息 进行设定

1. 支持实体对的最大长度

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u initial.py -max\_entities\_len=32 > logs/initial.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_train.py -weight='relation' -max\_entities\_len=32 > logs/train.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_test.py -weight='relation' -max\_entities\_len=32> logs/test.log 2>&1 &

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master# nohup python -u model\_predict.py -weight='relation' -max\_entities\_len=32> logs/ predict.log 2>&1 &

注意：max\_entities\_len只有在weight='relation'时才有效；max\_entities\_le的长度设定可以参考4.1章节，第一步中数据集的基本信息进行设定

1. 设置深度循环神经网络的层数

(tensorflow\_py37\_gpu) root@gpu:/sk/relation\_extract/Relationship-extraction-master#nohup python -u model\_train.py -num\_layers=2 > logs/train.log 2>&1 &

注意：如果训练时设置了层数，那么执行model\_test，model\_predict时也必须设置同样的层数即：

nohup python -u model\_test.py -num\_layers=2 > logs/test.log 2>&1 &

nohup python -u model\_predict.py -num\_layers=2 > logs/predict.log 2>&1 &

