

“刘知远实验室”的关系抽取实验复现

原创

小时不识月123 于 2018-09-27 17:04:05 发布 6869 收藏 13

分类专栏: [机器学习工程思考](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: <https://blog.csdn.net/leitouguan8655/article/details/82869881>

版权



[机器学习工程思考](#) 专栏收录该内容

11 篇文章 1 订阅

订阅专栏

最近在做文本的关系抽取, 看了一篇论文 (NRE论文总结: [Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification](#)):

没有找到论文源码, 我就从[github](#)上找到这份脚本 (来自“刘知远实验室”):

[thunlp/OpenNRE](#)

[thunlp/OpenNRE](#)

下面便开始了我的艰辛的调试脚本之路

1. 下数据, 解压时候遇到了问题

[tar](#)解压Unexpected EOF in archive错误

2. tensorflow-gpu==1.4.1、anaconda2(python2)、linux环境下跑脚本

原脚本要求: [TensorFlow \(>=1.4.1\)](#), 所以我就想装1.4.1版本的tensorflow

[tensorflow](#)指定版本的安装及升级到最新版

```
pip install tensorflow-gpu==1.4.1
```

装完之后想使用, 发现报错了:

```
libcudnn.so.6:cannot open shared object file:No such file or directory
```

然后我在anaconda的.bashrc文件里加了一行配置环境 (cudnn-6.0), 完美解决。

3. tensorflow-cpu==1.2、anaconda3(python3)、window环境下跑脚本

我还尝试了用tensorflow1.2+anaconda3(python3)版本的跑, 报错如下:

报错1:

```
AttributeError: module 'tensorflow' has no attribute 'AUTO_REUSE'
```

原因: tensorflow版本太低了, 'AUTO_REUSE'是1.4.0的语法。

修改: 将 `encoder.py` 和 `selector.py` 里的 `reuse=tf.AUTO_REUSE` 都修改为 `reuse=True`。

没有跑通, 就放假了, 回老家啦~

国庆回来, 刚想拾起上面的脚本仔细研读, 发现作者更新了: 新的脚本地址, 我决定看新版本的脚本, 因为代码量少了很多, 很多, 很多, 如果你在看我写的学习笔记, 建议上面的那些文字都不要看啦, 只要看下面的就好啦~~

环境: anaconda2+tensorflow1.5+gpu+CUDA8.0

1.数据转化为json时, 作者给的运行语句有些问题, 应该用下面的语句:

```
python protobuf2json.py path (给一个路径)
```

2.执行: `python train_demo.py nyt pcnn att` 时:

改报错1: `'cmp' is an invalid keyword argument for this function`

解决办法:

将 `~\OpenNRE-master\nrekit\data_loader.py` 中第260行的:

```
self.ori_data.sort(cmp=compare_by_entities_and_relations)
```

改为: [cmp报错的资料](#)

```
from functools import cmp_to_key#这句放在脚本文件最开始的地方
self.ori_data.sort(key=cmp_to_key(compare_by_entities_and_relations))
```

改报错2: `No such file or directory: '_processed_data\data\nyt\~`

**解决办法: **创建路径 `~\OpenNRE-master_processed_data\data\nyt`

改报错3: `'dict' object has no attribute 'iteritems'`

**解决办法: **将 `~\OpenNRE-master\nrekit\data_loader.py` 中第426、433行的 `iteritems` 变为 `items`。

3.然后脚本就跑通啦

结果: PCNN+ATTENTION时, “刘知远实验室”给出的AUC是0.3408, 我实验跑出来的结果是0.1514, AUC越接近1说明模型衡量分类的能力越好, 也就是说关系抽取在目前来说, 做出来的效果都很差。

先这样吧, 脚本和论文都没时间研读啦, 我要离开一段时间了, 祝我一切都好!