博客园

首页　　新闻　　博问　　专区　　闪存　　班级　　　代码改变世界　　　　　注册　　登录

天花板的学习笔记

种一棵树最好的时间是十年前,其次是现在。

博客园　　　首页　　　新随笔　　　联系　　　订阅　　　管理

随笔 - 6　文章 - 0　评论 - 3　阅读 - 9927

配置、使用transformers包

‹		2021年4月				›
日	一	二	三	四	五	六
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	1
2	3	4	5	6	7	8

搜索

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

随笔档案

2019年11月(2)
2019年10月(4)

最新评论

1. Re:配置、使用transformers包
你好，为什么我的测试代码运行后，提示
ModuleNotFoundError: No module named
'fused_layer_norm_cuda'呢?
　　　　　　　　--2014清风2014
2. Re:配置、使用transformers包
66666666666666666666666666666666
66666666666666666666666666666666
66666666666666666666
　　　　　　　--台北蒋介石
3. Re:配置、使用transformers包
嗯，虽然比较简单，但是可行
　　　　　　　--任小牛

配置、使用transformers包

一、transformers

transformers包又名pytorch-transformers或者pytorch-pretrained-bert。它提供了一些列的STOA模型的实现，包括(Bert、XLNet、RoBERTa等)。下面介绍该包的使用方法：

1、如何安装

transformers的安装十分简单，通过pip命令即可

```
pip install transformers
```

也可通过其他方式来安装，具体可以参考：https://github.com/huggingface/transformers

2、如何使用

使用transformers前需要下载好pytorch(版本>=1.0)或者tensorflow2.0。下面以pytorch为例，来演示使用方法

1、若要导入所有包可以输入：

```
import torch
from transformers import *
```

2、若要导入指定的包可以输入：

```
import torch
from transformers import BertModel
```

3、加载预训练权重和词表

```
UNCASED = './bert-base-uncased'
bert = BertModel.from_pretrained(UNCASED)
```

注意：**加载预训练权重时需要下载好预训练的权重文件，一般来说，当缓存文件中没有所需文件时(第一次使用)，只要网络没有问题，就会自动下载。当网络出现问题的时候，就需要手动下载预训练权重了。**

当缓存中不存在所需文件时，一般会出现提示：

bert-base-uncased-pytorch_model.bin not found in cache

二、手动下载、本地加载预训练权重

在使用transformers的时候，由于Bert、XLNet的文件都在AWS上存储，transformers的默认下载地址指向的是AWS，因此在国内下载速度非常慢。需要我们自己手动下载。

1、下载.txt、.json、.bin文件到本地

以Bert为例，相关的.bin文件(预训练权重)下载地址如下所示：

```
BERT_PRETRAINED_MODEL_ARCHIVE_MAP = {
    'bert-base-uncased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
base-uncased-pytorch_model.bin",
    'bert-large-uncased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
large-uncased-pytorch_model.bin",
    'bert-base-cased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-
cased-pytorch_model.bin",
    'bert-large-cased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
large-cased-pytorch_model.bin",
    'bert-base-multilingual-uncased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-multilingual-uncased-
pytorch_model.bin",
    'bert-base-multilingual-cased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-multilingual-cased-
pytorch_model.bin",
    'bert-base-chinese': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
base-chinese-pytorch_model.bin",
    'bert-base-german-cased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-german-cased-
pytorch_model.bin",
    'bert-large-uncased-whole-word-masking':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-uncased-whole-word-
masking-pytorch_model.bin",
    'bert-large-cased-whole-word-masking':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-cased-whole-word-
masking-pytorch_model.bin",
    'bert-large-uncased-whole-word-masking-finetuned-squad':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-uncased-whole-word-
masking-finetuned-squad-pytorch_model.bin",
    'bert-large-cased-whole-word-masking-finetuned-squad':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-cased-whole-word-
masking-finetuned-squad-pytorch_model.bin",
    'bert-base-cased-finetuned-mrpc':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-cased-finetuned-mrpc-
pytorch_model.bin",
    'bert-base-german-dbmdz-cased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-german-dbmdz-cased-
pytorch_model.bin",
    'bert-base-german-dbmdz-uncased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-german-dbmdz-uncased-
pytorch_model.bin",
}
```

若需要下载.json文件，则下载地址为：

```
BERT_PRETRAINED_CONFIG_ARCHIVE_MAP = {
    'bert-base-uncased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
base-uncased-config.json",
    'bert-large-uncased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
large-uncased-config.json",
    'bert-base-cased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-
cased-config.json",
    'bert-large-cased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
large-cased-config.json",
    'bert-base-multilingual-uncased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-multilingual-uncased-
config.json",
    'bert-base-multilingual-cased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-multilingual-cased-
config.json",
    'bert-base-chinese': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
base-chinese-config.json",
    'bert-base-german-cased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-german-cased-
config.json",
    'bert-large-uncased-whole-word-masking':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-uncased-whole-word-
masking-config.json",
    'bert-large-cased-whole-word-masking':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-cased-whole-word-
masking-config.json",
    'bert-large-uncased-whole-word-masking-finetuned-squad':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-uncased-whole-word-
masking-finetuned-squad-config.json",
    'bert-large-cased-whole-word-masking-finetuned-squad':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-cased-whole-word-
masking-finetuned-squad-config.json",
    'bert-base-cased-finetuned-mrpc':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-cased-finetuned-mrpc-
config.json",
    'bert-base-german-dbmdz-cased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-german-dbmdz-cased-
config.json",
    'bert-base-german-dbmdz-uncased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-german-dbmdz-uncased-
config.json",
}
```

.txt相关文件(词表文件)下载地址如下：

```
PRETRAINED_VOCAB_FILES_MAP = {
    'vocab_file':
    {
        'bert-base-uncased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
base-uncased-vocab.txt",
        'bert-large-uncased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-uncased-vocab.txt",
        'bert-base-cased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
base-cased-vocab.txt",
        'bert-large-cased': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
large-cased-vocab.txt",
        'bert-base-multilingual-uncased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-multilingual-uncased-
vocab.txt",
        'bert-base-multilingual-cased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-multilingual-cased-
vocab.txt",
        'bert-base-chinese': "https://s3.amazonaws.com/models.huggingface.co/bert/bert-
base-chinese-vocab.txt",
        'bert-base-german-cased': "https://int-deepset-models-bert.s3.eu-central-
1.amazonaws.com/pytorch/bert-base-german-cased-vocab.txt",
        'bert-large-uncased-whole-word-masking':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-uncased-whole-word-
masking-vocab.txt",
        'bert-large-cased-whole-word-masking':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-cased-whole-word-
masking-vocab.txt",
        'bert-large-uncased-whole-word-masking-finetuned-squad':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-uncased-whole-word-
masking-finetuned-squad-vocab.txt",
        'bert-large-cased-whole-word-masking-finetuned-squad':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-large-cased-whole-word-
masking-finetuned-squad-vocab.txt",
        'bert-base-cased-finetuned-mrpc':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-cased-finetuned-mrpc-
vocab.txt",
        'bert-base-german-dbmdz-cased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-german-dbmdz-cased-
vocab.txt",
        'bert-base-german-dbmdz-uncased':
"https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-german-dbmdz-uncased-
```

```
    vocab.txt",
    }
}
```

本地使用预训练的权重文件

假设我们下载好了如下三个文件：

bert-base-uncased-config.json	2019/11/27 16:23	JSON 文件	1 KB
bert-base-uncased-pytorch_model.bin	2019/11/26 17:00	BIN 文件	430,150 KB
bert-base-uncased-vocab.txt	2019/11/27 16:23	文本文档	227 KB

首先我们建立一个文件夹，命名为bert-base-uncased，然后将这个三个文件放入这个文件夹，并且对文件进行重命名，重命名时将bert-base-uncased-去除即可。

假设我们训练文件夹名字为 train.py，我们需要将上面的bert-base-uncased文件夹放到与train.py同级的目录下面。
若不改名以及调整文件夹位置将会出现：
vocab.txt not found；pytorch_model.bin not found；Model name 'xxx/pytorch_model.bin ' was not found in model name list等错误。

之后使用下面的代码进行测试即可：

```
UNCASED = './bert-base-uncased'
VOCAB = 'vocab.txt'
tokenizer=BertTokenizer.from_pretrained(os.path.join(UNCASED,VOCAB))
bert = BertModel.from_pretrained(UNCASED)
```

好文要顶　关注我　收藏该文

« 上一篇： Centos安装Nvidia驱动、Cuda、Cudnn、TF2.0

posted @ 2019-11-28 10:56 天花板掉啦a 阅读(8160) 评论(3) 编辑 收藏

刷新评论　刷新页面　返回顶部

登录后才能查看或发表评论，立即 登录 或者 逛逛 博客园首页

【推荐】世界读书日活动--记录阅读之路，影响同行之人
【推荐】全球最大规模开发者调查启动--你的声音，值得让世界听见！
【推荐】大型组态、工控、仿真、CAD\GIS 50万行VC++源码免费下载！
【推荐】华为终端云服务创新实验室--神奇的室内地磁定位技术
【推荐】限时秒杀！国云大数据魔镜，企业级云分析平台

园子动态：
· 致园友们的一封检讨书：都是我们的错
· 数据库实例 CPU 100% 引发全站故障
· 发起一个开源项目：博客引擎 fluss