> 本文档记录了文本分类/情感分析英文版模型接口 v2 的相关内容
>
> Author：成臻

# 文件目录

- `pkgs`：包含模型调用所用到接口函数
  - `data_preprocess.py`：输入数据预处理的相关函数
  - `wrapper.py`：模型启动调用的相关函数
  - `interface.py`：对外接口
- `models`：目前只包含英文版句子级情感 5 分类的测试模型，按照模型类别分别存储在四个文件中
  - `cnn`
  - `rnn`
  - `rcnn`
  - `clstm`

# 接口调用

- 模型输入
  - 需要进行文本分类/情感分析的句子
  - 数据类型：`str` in `python3`
- 模型输出：
  - 句子的文本分类/情感分析标签
  - 目前仅以 5 分类的情感分析作为输出，即 0～4 的标签，含义如下
    - 0：很差
    - 1：差
    - 2：中性
    - 3：好
    - 4：很好
  - 数据类型：`list of int` in `python3`，list 只包含一个整数值，即标签
- 后期模型扩展：之后模型输入不会变化，主要会改变为三种输出类型，依赖于 `数据标注方案v3` 和 `数据标注方案V3--关于文本分类标准的补充说明`，具体以文档为准（存在负值情况下模型输出会产生自动偏移，如 -2 更改为 0，-1 更改为 1）

# 环境要求

- python3.5+
- tensorflow 1.4+
- CUDA 8.0
- numpy
- gensim

# 调用方式

目前仅通过 `console` 进行测试，测试方式如下：

```
python3 interface.py [text|sen3|sen5]
```

有且仅有一个参数来指定任务：

- text：调用文本分类模型
- sen3：调用 3 级情感分类模型
- sen5：调用 5 级情感分类模型

该命令直接启动模型，并处于等待命令行输入状态。

每次输入一个句子之后会将模型输出直接打印到命令行显示，具体输出在 `wrapper.py` 的 `text_classification` 可以找到，如需更改输入输出方式，更改以下代码：

```
while True:
    sentence = input("please input a sentence in English, MAX LENGTH=50\n")
    formated_sentence_idx = formate_str(sentence, vocab)
    pred_y, acc = sess.run([predictions, accuracy], {
        input_x: formated_sentence_idx,
        input_y: [[0, 0, 0, 0 ,0]],
        dropout_keep_prob: 1.0
    })
    print(pred_y)
```