**TensorFlow｜7｜混合数据集**

1. 基本环境

|  |  |
| --- | --- |
| 程序名称 | mixData.py |
| 硬件环境 | MacOSX |
| 软件环境 | Python3, Spyder, Anaconda |

1. 核心问题：将两类数据集混合，然后在新数据集上测试网络的表现
2. 解决过程：

由于之前测试用的数据集大小为10000，前5000为负面评价，后5000为正面评价。取4000到6000为测试集，其余为训练集。为了令后期程序尽可能不该动，这里依旧按照这个方式构造新的混合数据集。

* 导入服饰类和水果类数据集
* 对于存储数据集的向量fileTrainRead，首先进行2500次循环，每一次输入一句服饰类负面评价、一句水果类负面评价；再进行2500次循环，每一次输入一句服饰类正面评价、一句水果类正面评价。这样就构成了一个新的数据集

1. 代码：

clothfilePath='/Users/xuyizhou/Desktop/clothing.txt'

fileDonePath ='/Users/xuyizhou/Desktop/corpus.txt'

#导入服饰类数据集

clothfileTrainRead = []

with open(clothfilePath) as clothfileTrainRaw:

for line in clothfileTrainRaw:

clothfileTrainRead.append(line)

#导入水果类数据集

fruitfilePath='/Users/xuyizhou/Desktop/fruit.txt'

fruitfileTrainRead = []

with open(fruitfilePath) as fruitfileTrainRaw:

for line in fruitfileTrainRaw:

fruitfileTrainRead.append(line)

#构建新的数据集

fileTrainRead=[]

for i in range(2500):

fileTrainRead.append(clothfileTrainRead[i])

fileTrainRead.append(fruitfileTrainRead[i])

for i in range(2500):

fileTrainRead.append(clothfileTrainRead[5000+i])

fileTrainRead.append(fruitfileTrainRead[5000+i])

with open(fileDonePath,'w') as fW:

for i in range(len(fileTrainRead)):

fW.write(str(fileTrainRead[i]))

1. 结果：

