有时候一个csv文件中有数字和非数字数据这时候如果你以数字的方式来读取文件，使用numpy的loadtxt方法会立马崩溃。

此时可以使用numpy的**genfromtxt(filename,** dtype=int, delimiter=',',skipheader=1)来加载，加载后非数字数据会显示nan

这时候可以自定义一个方法来替换nan

|  |
| --- |
| def replace\_nan(ndarr): # 自己封装的方法，可以替换任意数组中的nan  # 处理nan，将其替换为这一列所有非nan是数值的均值  for i in range(ndarr.shape[1]):  tmp\_col = ndarr[:, i]  nan\_count = np.count\_nonzero(tmp\_col != tmp\_col)  if nan\_count != 0: # 说明这一列有nan  # 将这一列的非nan取出来计算平均值  tmp\_non\_nan\_col = tmp\_col[tmp\_col == tmp\_col]  avg = tmp\_non\_nan\_col.mean()  tmp\_col[tmp\_col != tmp\_col] = avg  return ndarr |

具体代码

|  |
| --- |
| import numpy as np  t\_uk = np.genfromtxt("./uk\_videos.csv", dtype=int, delimiter=",",skip\_header=1)  # 使用散点图显示uk数据这like和comments的关系 t\_uk = replace\_nan(t\_uk) np.savetxt("uk\_data.csv", t\_uk, delimiter=",", fmt="%s") #保存替换后的数据，实现清除非数字数据的目的  # 如果需要表头，可以用excel打开csv文件在第一行插入表头数据 |