

python , numpy , pandas数据处理之小技巧



Edward.Fu · 2 年前

今日是端午4天假期的最后一天，自己的2台电脑一台在处理数据，另一台电源适配器也坏掉了。感觉闲来无事，在此就简单聊聊numpy和pandas的各种冷知识、小技巧 and 疑难杂症。个人觉得关于numpy和pandas的坑太大，不好填，在此就不系统的介绍了，各位可以买些这块的资料先自己看看。

这里就着重聊聊一些使用过程中常用到但教科书里找不着的问题，省的各位朋友还跑去stackoverflow找答案。

问题一、dataframe里面.values , .iloc , .ix , .loc的区别？

只有values是将原本dataframe数据强制转化为numpy格式的数据来索引，其他3个都是对dataframe本身的数据来索引，其中iloc是对基于values的位置来索引调用，loc是对index和columns的位置来索引，而ix则是先用loc的方式来索引，索引失败就转成iloc的方式；

问题二、可否有两层，或2层以上的columns或index？有的话如何索引？

可以，索引的话如果用loc或ix，则默认是用第一层的index或columns，最简单的方式是类似于这样：

```
example.loc[index1, columns1].loc[index2, columns2]
```

问题三、list, dict, numpy.ndarray, dataframe数据格式如何转换？

1. list转化为numpy.ndarray：

```
np.array(example)
```

2. numpy.ndarray转化为list：

```
list(example)
```

3. dict转化为dataframe:

