

首页

博客

学院

下载

论坛

APP

问答

商城

活动

VIP会员

续费9折

专题

招聘

ITeye

GitChat

图文课

原

Pandas：分组级的运算和转换--transform和apply

2018年04月19日 22:03:48 BQW_ 阅读量 3005

csdn 版权声明：本文为博主原创文章，未经博主允许不得转载。 <https://blog.csdn.net/bqw18744018044/article/details/80011733>

```
1 import numpy as np
2 import pandas as pd
3 from pandas import Series,DataFrame
```

一、介绍

聚集函数只是将一组值转换为一个标量值，其实还可以将更多类型的函数应用到分组上，下面的transform和apply就提供这样的功能。

二、transform

transform中传入的函数只能返回两种结果，可以广播的标量值或者与分组大小相同的结果数组。

```
1 df = DataFrame({'类别':['水果','水果','水果','蔬菜','蔬菜','肉类','肉类'],
2                 '产地':['美国','中国','中国','中国','新西兰','新西兰','美国'],
3                 '水果':['苹果','梨','草莓','番茄','黄瓜','羊肉','牛肉'],
4                 '数量':[5,5,9,3,2,10,8],
5                 '价格':[5,5,10,3,3,13,20]})
6 print(df)
```

	产地	价格	数量	水果	类别
0	美国	5	5	苹果	水果
1	中国	5	5	梨	水果
2	中国	10	9	草莓	水果
3	中国	3	3	番茄	蔬菜
4	新西兰	3	2	黄瓜	蔬菜
5	新西兰	13	10	羊肉	肉类
6	美国	20	8	牛肉	肉类

1.对分组求均值，然后把这个均值赋值给整个组（可广播的标量值）

```
1 print(df.groupby('类别').transform(np.mean))
```

	价格	数量
0	6.666667	6.333333
1	6.666667	6.333333
2	6.666667	6.333333
3	3.000000	2.500000
4	3.000000	2.500000
5	16.500000	9.000000
6	16.500000	9.000000

2.使用transform实现去均值（返回与分组大小相同的结果数组）

```
1 demean = lambda arr:arr-arr.mean()
2 print(df.groupby('类别').transform(demean))
```

	价格	数量
0	-1.666667	-1.333333
1	-1.666667	-1.333333
2	3.333333	2.666667
3	0.000000	0.500000
4	0.000000	-0.500000
5	-3.500000	1.000000
6	3.500000	-1.000000

三、apply

apply会将分组传递给函数，然后尝试将函数的返回结果组合起来。



```
1 def get_rows(df,n=1): # 返回分组的前n行数据
2     return df.iloc[:n,:]
3 print(df.groupby('类别').apply(get_rows))
```

```
1         产地   价格   数量   水果   类别
2 类别
3 水果 0    美国    5    5   苹果   水果
4 肉类 5   新西兰   13   10   羊肉   肉类
5 蔬菜 3    中国    3    3   番茄   蔬菜
```

为apply的函数传递参数

```
1 print(df.groupby('类别').apply(get_rows,n=2))
```

```
1         产地   价格   数量   水果   类别
2 类别
3 水果 0    美国    5    5   苹果   水果
4     1    中国    5    5    梨   水果
5 肉类 5   新西兰   13   10   羊肉   肉类
6     6    美国   20    8   牛肉   肉类
7 蔬菜 3    中国    3    3   番茄   蔬菜
8     4   新西兰    3    2   黄瓜   蔬菜
```

股神老徐狱中曝出庄家洗盘规律，牢记这3点，看完恍然大悟！

股管家 · 顶新

 想对作者说点什么

pandas apply vs agg vs transform

DataPreparationimportpandasaspdfimportnumpyasnpdf=pd.DataFrame({'A':[1,1,2,2],'B':[1,2,3,4],'C':np.random.randn(4)})

博文 来自: [fanfanyuzhui的专栏](#) 阅读数 2219

Python3pandas库transform用法

```
importpandasaspdfimportnumpyasnpA=np.array([[1,2,3,4,5],[2,1,1,2,2],[1,2,3,4,5],[2,1,1,2,2],[1,2,3,4,5]])
```

博文 来自: [Young_618](#) 阅读数 7079

通俗地讲清楚fit_transform()和transform()的区别

网上抄来抄去都是一个意思，fit_transform是fit和transform的组合。我们知道fit (x, y) 在新手入门的例子中比较...

博文 来自: [Chi Yu's Blog](#) 阅读数 3万+

Pandas对DataFrame单列/多列进行运算 (map, apply, transform, agg)

1.单列运算在Pandas中，DataFrame的一列就是一个Series,可以通过map来对一列进行操作： df['col2']=df['col1']...

博文 来自: [zwhooo的博客](#) 阅读数 3万+

入门级Python练手项目，有哪些值得推荐？附学习图谱

这几个练手项目，解决了很多常见问题

广告 关闭

Pandas玩转数据 (六) -- 通过apply对数据进行处理

```
importpandasaspdfimportnumpyasnpfrompandasimportSeries,DataFrame#打开一个csv文件对其中一列分成几...
```

博文 来自: [越看越喜欢啊](#) 阅读数 763

Python3 pandas库transform用法

```
importpandasaspdfimportnumpyasnpA=np.array([[1,2,3,4,5],[2,1,1,2,2],[1,2,3,4,5],[2,1,1,2,2],[1,2,3,4,5]])
```

博文 来自: [qq_41231470的博客](#) 阅读数 445

pandas小记： pandas数据规整化-分组合并及重塑

http://blog.csdn.net/pipisorry/article/details/39506169数据分析和建模方面的大量编程工作都是用在数据准备上...

博文 来自: [皮皮blog](#) 阅读数 2万+

pandas常用函数

```
importnumpyasnpimportpandasaspdfimportmatplotlib.pyplotasplt-----numpy-----...
```

博文 来自: [Zxdon的专栏](#) 阅读数 1万+

pandas使用笔记

```
DataFramedates=pd.date_range('20160728',periods=6)#创建固定频度的时间序列df=pd.DataFrame(np.random.randn(6,2))
```

博文 来自: [小母的世界](#) 阅读数 1万+

高精度3d打印机 快速成型 极光尔沃3D打印机厂家大型3d打印机

1688热销

 1

















