新殖電

# 【转】PANDAS 数据合并与重塑 (concat篇)

订阅

庆系

## 转自:

中容回

自灾

http://blog.csdn.net/stevenkwong/article/details/52528616

### 1 concat

#### 参数说明

objs: series, dataframe或者是panel构成的序列Isit

axis: 需要合并链接的轴, 0是行, 1是列join: 连接的方式 inner, 或者outer

其他一些参数不常用,用的时候再补上说明。

#### 1.1 相同字段的表首尾相接

		df1					Result		
	Α	В	С	D					
0	A0	В0	00	D0		Α	В	С	D
1	A1	B1	C1	D1	0	A0	В0	ω	D0
2	A2	B2	C2	D2	1	Al	B1	C1	D1
3	A3	В3	C3	D3	2	A2	B2	(2	D2
		df2							
	Α	В	С	D	3	A3	В3	СЗ	D3
4	A4	B4	C4	D4	4	A4	B4	C4	D4
5	A5	B5	C5	D5	5	A5	B5	C5	D5
6	Аб	В6	C6	D6	6	A6	В6	C6	D6
7	A7	B7	C7	D7	7	A7	В7	C7	D7
		df3							
	Α	В	С	D	8	A8	B8	C8	DB
8	A8	B8	C8	DB	9	A9	B9	C9	D9
9	A9	B9	C9	D9	10	A10	B10	C10	D10
10	A10	B10	C10	D10	11	A11	B11	C11	D11
11	A11	B11	C11	D11					

```
# 现将表构成list, 然后在作为concat的输入
In [4]: frames = [df1, df2, df3]
In [5]: result = pd.concat(frames)
```

要在相接的时候在加上一个层次的key来识别数据源自于哪张表,可以增加key参数

```
In [6]: result = pd.concat(frames, keys=['x', 'y', 'z'])
```

效果如下

公告

昵称: stAr\_1 园龄: 2年3个月

27

搜索

25

1

我的标签

Tree(33)

LinkedList(19

Math(19)

HashTable(1

String(7)

动态规划(5)

Two Points(4

HashMap(3)

Array(2)

dfs(2)

		df1					Res	sult		
	Α	В	С	D						
0	A0	В0	α	D0			Α	В	С	D
1	A1	B1	C1	D1	х	0	A0	В0	α	D0
2	A2	B2	C2	D2	×	1	Al	B1	а	D1
3	A3	В3	C3	D3	×	2	A2	B2	2	D2
		df2			_ ^		~	LLE	4	L/Z
	Α	В	С	D	х	3	A3	B3	СЗ	D3
4	A4	B4	C4	D4	у	4	A4	B4	C4	D4
5	A5	B5	C5	D5	у	5	A5	B5	C5	D5
6	Аб	В6	C6	D6	у	6	Аб	В6	Cl6	D6
7	A7	B7	C7	D7	у	7	A7	B7	C7	D7
		df3								
	Α	В	С	D	Z	8	AB	B8	CB	D8
8	A8	B8	C8	DB	z	9	A9	B9	C9	D9
9	A9	B9	C9	D9	z	10	A10	B10	C10	D10
10	A10	B10	C10	D10	z	11	A11	B11	C11	D11
11	A11	B11	C11	D11						

#### 1.2 横向表拼接 (行对齐)

#### 1.2.1 axis

当axis = 1的时候, concat就是行对齐, 然后将不同列名称的两张表合并

In [9]: result = pd.concat([df1, df4], axis=1)

		df1				df	4					Res	sult			
										Α	В	С	D	В	D	F
	Α	В	С	D		В	D	F	0	A0	В0	co	D0	NaN	NaN	NaN
0	A0	B0	co	D0	2	B2	D2	F2	1	A1	B1	C1	D1	NaN	NaN	NaN
1	A1	B1	Cl	D1	3	В3	D3	F3	2	A2	B2	C2	D2	B2	D2	F2
2	A2	B2	C2	D2	6	B6	D6	F6	3	A3	В3	СЗ	D3	В3	D3	F3
3	A3	В3	C3	D3	7	B7	D7	F7	6	NaN	NaN	NaN	NaN	B6	D6	F6
									7	NaN	NaN	NaN	NaN	B7	D7	F7

#### 1.2.2 join

加上join参数的属性,如果为'inner'得到的是两表的交集,如果是outer,得到的是两表的并集。

In [10]: result = pd.concat([df1, df4], axis=1, join='inner')

		df1				df	4				A2 B2 C2 D2 B2 D2 F2					
	Α	В	С	D		В	D	F								
0	A0	BO	α	D0	2	B2	D2	F2		Α	В	С	D	В	D	F
1	A1	B1	C1	D1	3	В3	D3	F3	2	A2	B2	C2	D2	B2	D2	F2
2	A2	B2	C2	D2	6	B6	D6	F6	3	A3	В3	СЗ	D3	В3	D3	F3
3	A3	В3	C3	D3	7	B7	D7	F7								

#### 1.2.3 join\_axes

如果有join\_axes的参数传入,可以指定根据那个轴来对齐数据 例如根据df1表对齐数据,就会保留指定的df1表的轴,然后将df4的表与之拼接

In [11]: result = pd.concat([df1, df4], axis=1, join\_axes=[df1.index])

		df1				df	4									
	Α	В	С	D		В	D	F		Α	В	С	D	В	D	F
0	A0	В0	œ	D0	2	B2	D2	F2	0	A0	В0	00	D0	NaN	NaN	NaN
1	A1	B1	C1	D1	3	В3	D3	F3	1	A1	B1	C1	D1	NaN	NaN	NaN
2	A2	B2	C2	D2	6	B6	D6	F6	2	A2	B2	C2	D2	B2	D2	F2
3	A3	В3	C3	D3	7	B7	D7	F7	3	A3	В3	СЗ	D3	В3	D3	F3

### 1.3 append

append是series和dataframe的方法,使用它就是默认沿着列进行凭借 (axis = 0, 列对齐)

In [12]: result = dfl.append(df2)

更多

随笔档案 2019年4月 (2 2018年9月 (2 2018年8月 (4 2018年7月 (2 2018年6月 (1 2018年5月 (1 2018年4月 (9 2018年3月 (1 2018年2月 (4 2018年1月 (6 2017年12月 2017年11月 2017年10月 2017年9月 (1 2017年8月 (2 2017年7月 (1

2017年6月 (4

2017年5月 (5

阅读排行榜

1. idea中mav

2. zotero使用

3. LR为什么用 为什么是log振

4. windows ₹ and "python 信息(3257)

5. positive-u (2193)

		df1				Result  A B C D  0 A0 B0 C0 D0  1 A1 B1 C1 D1  2 A2 B2 C2 D2  3 A3 B3 C3 D3  4 A4 B4 C4 D4				
	Α	В	С	D		А	В	С	D	
0	A0	В0	α	D0						
1	Al	B1	C1	D1	0	AD.	B0	ω	DO	
2	A2	B2	C2	D2	1	A1	B1	C1	D1	
3	A3	В3	C3	D3	2	A2	B2	C2	D2	
		df2			3	A3	В3	СЗ	D3	
	Α	В	С	D	4	A4	B4	C4	D4	
4	A4	B4	C4	D4	5	A5	B5	C5	D5	
5	A5	B5	C5	D5		1.0	D.C.	~	D.C.	
6	Аб	В6	C6	D6	6	Аб	B6	O6	D6	
7	A7	В7	C7	D7	7	A7	B7	C7	D7	

#### 1.4 无视index的concat

如果两个表的index都没有实际含义,使用ignore\_index参数,置true,合并的两个表就睡根据列字段对齐,然后合并。最后再重新整理一个新的index。

		df1					Res	sult		
	Α	В	С	D		Α	В	С	D	F
0	A0	В0	a	D0						
1	Al	B1	C	1 D1	0	A0	B0	00	D0	NaN
2	A2	B2	C	2 D2	1	A1	B1	C1	D1	NaN
3	A3	В3	C	3 D3	2	A2	B2	C2	D2	NaN
		df4			3	A3	В3	СЗ	D3	NaN
	В		D	F	4	NaN	B2	NaN	D2	F2
	_	B2	D2	F2	5	NaN	В3	NaN	D3	F3
3	3	B3	D3	F3	6	NaN	B6	NaN	D6	F6
(	5	B6	D6	F6						
7	7	B7	D7	F7	7	NaN	B7	NaN	D7	F7
					J					

#### 1.5 合并的同时增加区分数据组的键

前面提到的keys参数可以用来给合并后的表增加key来区分不同的表数据来源

#### 1.5.1 可以直接用key参数实现

In [27]: result = pd.concat(frames, keys=['x', 'y', 'z']) df1 Result C D D D0  $\alpha$ D1 Al В1 Cl A2 В2 C2 D2 D1 АЗ ВЗ СЗ D3 D2 A2 D3 В C D D4 D4 A4 C4 D5 A5 B5 D5 D6 A7 В7 C7 D7 D7 D8 В C D D9 A8 B8 C8 DB A10 B10 D10 Α9 C9 D9 A10 C10 11 D11

#### 1.5.2 传入字典来增加分组键

A11

B11

C11 D11

```
In [28]: pieces = {'x': df1, 'y': df2, 'z': df3}
In [29]: result = pd.concat(pieces)
```

推荐排行榜

1. [leetcode]

题目: 跳数游

2. idea中surr

3. idea中mav

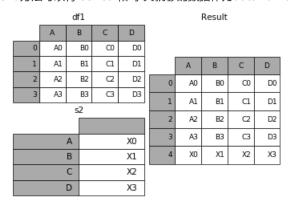
4. 740. Dele

5. zotero使用

		df1					Res	sult		
	Α	В	С	D						
0	A0	B0	co	D0			Α	В	С	D
1	A1	B1	C1	D1	х	0	A0	B0	α	D0
2	A2	B2	C2	D2	×	1	Al	B1	а	D1
3	A3	В3	C3	D3	×	2	A2	B2	(2	D2
		df2							_	
	Α	В	С	D	х	3	A3	B3	СЗ	D3
4	A4	B4	C4	D4	у	4	A4	B4	C4	D4
5	A5	B5	C5	D5	у	5	A5	B5	C5	D5
6	A6	В6	C6	D6	у	6	Аб	B6	Cl6	D6
7	A7	B7	C7	D7	у	7	A7	B7	C7	D7
		df3								
	Α	В	С	D	Z	8	AB	B8	CB	D8
8	A8	B8	C8	DB	z	9	A9	B9	(9	D9
9	A9	B9	C9	D9	z	10	A10	B10	C10	D10
10	A10	B10	C10	D10	z	11	Al1	B11	C11	D11
11	A11	B11	C11	D11						

#### 1.6 在dataframe中加入新的行

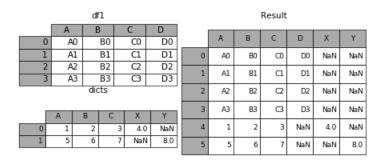
append方法可以将 series 和 字典就够的数据作为dataframe的新一行插入。



In [34]: s2 = pd.Series(['X0', 'X1', 'X2', 'X3'], index=['A', 'B', 'C', 'D'])
In [35]: result = df1.append(s2, ignore\_index=True)

#### 表格列字段不同的表合并

如果遇到两张表的列字段本来就不一样,但又想将两个表合并,其中无效的值用nan来表示。那么可以使用ignore\_index来实现。







0 0

« 上一篇: python对离散数据进行编码

» 下一篇: 朴素贝叶斯法

posted @ 2018-01-10 17:35 stAr\_1 阅读(456) 评论(0) 编辑 收藏