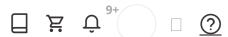


我的







pandas Split-Apply-Combine Strategy

menu_book	comput	<u>er</u> <u>F</u>
簡報閱讀	<u>範例與作</u>	<u>問題討論</u> <u>學習心得(完成)</u>
重要知識點		
認識 groupby		重要知識點
認識 Split-Apply- Combine 策略		介紹如何透過 pandas groupby 函式實現資料科學的 Split-Apply-Combine 策略
Groupby 針對多個欄位 做分析		認識 groupby
		在數據分析中時常會分析不同族群的資料,例如,

的邏輯(下圖 3)。

(表1)

學生分數資料(如表 1),你想分析男生與女生的各

科差異,前幾天有教到檢索可以將資料分成男生資

料與女生資料,在將各資料算平均值(如圖 2),在

這裡有一個函數 goupby 可以一行指令執行以上







運用 groupby 平均(圖3)

認識 Split-Apply-Combine 策略

以剛剛學生資料來分解一下 groupby 的邏輯過程

- Split:將大的數據集拆成可獨立計算的 小數據集(拆成男生、女生資料)
- Apply:獨立計算各個小數據集(成績取 平均)
- Combine:將小數據集運算結果合併

將 DataFrame 依照 A、B、C 拆成三個小數據集 [split],各自計算總合[Apply],合併結果輸出 [Combine]

 拆分成 A、B、C 小數據集的方法為 groupby

Groupby 針對多個欄位做分析

Groupby 也可以針對多個欄位做分析,例如,學 生成績資料多一欄位 c 班級(class), 想對班級以及 性別做分類,在 groupby 中加入兩個欄位名稱即 可(如下圖),此時 groupby 自動會生成多維度索 引(multiple index)

Groupby 針對欄位做多個分析







計算,在 groupby 後加入 agg() (如下圖),在 agg 中加入計算的邏輯(mean,std),此時 groupby 自 動會生成多維度欄位(multiple columns)

Groupby 同時針對多個欄位做多個 分析

- Groupby 也可以同時針對多個欄位做多個 **分析**,例如,學生成績資料,想針對性別、 班級做成績平均以及最高分的計算
- 合併了多欄位以及多分析

知識點回顧

- Groupby 可以拆成
 - Split:將大的數據集拆成可獨立計算的 小數據集
 - Apply:獨立計算各個小數據集
 - Combine:將小數據集運算結果合併
- Groupby 可以同時針對多個欄位做多個分 析

參考資料

groupby

網站:python/pandas數據挖掘(十四)groupby,聚合·分組級運算







Split-Apply-Combine Strategy for Data Mining

延伸閱讀

Pandas 分组 (GroupBy)

網站:<u>易百教程</u>

Pandas 的 groupby 語法

網站:Justim的喃喃自語

下一步:閱讀範例與完成作業