

題目：

FP-Growth Algorithm是一個找frequent pattern非常出名的演算法，比Apriori Algorithm快速。用Python將這個演算法實作出來是一個很好的練習使用Python程式的題目。在網路上可以找到FP-Growth Algorithm相關程式碼，希望同學在這個分組練習能夠自行完成這個演算法(當然可以參考別人的寫法)。

wiki: (https://en.wikipedia.org/wiki/Apriori_algorithm#cite_note-apriori-1)https://en.wikipedia.org/wiki/Association_rule_learning#cite_note-24 (https://en.wikipedia.org/wiki/Association_rule_learning#cite_note-24)
algorithm:https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.4436 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.4436)

簡要說明：

X=>Y稱為一個association rule，其中X稱為此規則的antecedent，Y稱為此規則的consequence。可延伸至多個items如3 items: X,Y=>Z。在這個作業我們在乎Association Rule:X=>Y的兩個值support與confidence
support(X,Y)=P(X,Y) = 多少筆記錄含有X與Y/總筆數
confidence(X=>Y)=P(X,Y)/P(X)

作業目標：

找到最多包含5 items的所有Association rule:
A=>B
A,B=>C
A=>B,C
A,B,C=>D
A,B=>C,D
A=>B,C,D
A,B,C,D=>E
A,B,C=>D,E
A,B=>C,D,E
A=>B,C,D,E
滿足support >= 0.1 (出現次數>=813)與confidence>=0.8。
資料檔:
附件資料檔含有8124筆記錄，每一筆資料具有某些屬性(以數字編號)。
資料檔內容:
1 3 9 13 23 25 34 36 38 40 52 54 59 63 67 76 85 86 90 93 98 107 113
2 3 9 14 23 26 34 36 39 40 52 55 59 63 67 76 85 86 90 93 99 108 114
...

在分組報告時，準備ppt(不含封面，>=3頁)並請強調下面幾點：

- 1.如何使用Python Collections加速計算(就是資料結構)。(1 ppt)
- 2.方便完成此分組作業的Python特性或基本模組。
- 3.重要程式片段說明。(1 ppt)
- 4.執行結果與整個計算時間。(1 ppt)

參考數據

- 1. 找到的frequent item set，大小由1到5，數目分別是


|L^1|=56
|L^2|=763
|L^3|=4593
|L^4|=16150
|L^5|=38800

- 2.找滿足條件的association rule，數目為

394175

- 3.整個處理應可在5秒內完成。

附件名
檔案大小

 mushroom.dat

557 KB

[📄 \(/api/uploads/reference/1041073/blob\)](/api/uploads/reference/1041073/blob)