概念分類［二］ 20160117

鑒於只用「單一概念」會讓分類結果過於主觀，造成偏差。
於是改用「概念群」的方式來分類，盼能讓分類結果更客觀。

實作方法：

A 定義概念

首先用人工訂出各個概念群，舉例如下：

```
Dic = {'脫俗':['脫俗','幽雅','清幽','優雅','氣質','心曠神怡'],\
       '輕鬆':['輕鬆','休閒','遊憩','悠閒','休憩','消遣','愜意','放鬆']}
```

如上，這裡定義了兩個概念群「脫俗」、「輕鬆」。概念群內有若干能代表該概念的詞語。此處概念群的標題「脫俗」、「輕鬆」僅為命名用，關鍵詞必須要與後方的內容(陣列內詞語)達到相似門檻，才會被認為是與此概念相似。

一、至於這些概念群是怎麼定義出來的？
(1) 觀看關鍵字結果，人工歸類
(2) 觀看關鍵字與哪些字詞相似度最高，進而定義概念
(3) ...
二、如何驗證概念群的定義好壞？
(1) 確認「概念內的相似度高」及「概念間的相似度低」
＊前兩點已寫好相關的工具，用此來定義出各個概念
（https://github.com/spicyscap/AttrClassification/blob/master/Classifaction_M2_DefineTools.ipynb）

B 概念分類

採用了雙重門檻「相似度」與「相似門檻」。一關鍵字要與某概念相符，必須滿足：
(1)關鍵字 與 概念內的詞語 的相似度 達到相似度門檻
(2)關鍵字 與 某概念內 達到相似度門檻的詞語 占該概念的總詞語數 達到一定的成數
(3)特殊情況：關鍵字若與某概念內的詞語達到完全相似，則會被直接歸類進去

舉例 ：
假設相似度 0.2 相似門檻 0.5。
有一概念 `'開心':['舒爽','快樂','好玩']`

有一詞語 ’ 爽 ’，想知道它是否可歸類於開心概念 ，於是開始比較 ：
發現，’ 爽 ’與'舒爽'相似度 0.4；與'快樂'相似度 0.3；與'好玩'相似度 0.1
發現，’ 爽 ’與'舒爽''快樂'都有達到相似度門檻(>0.2)，'好玩'則沒有。
也就是’ 爽 ’與'開心'概念中的 2／3 詞語達成相似，換算成小數約 0.67，已達到相似門檻
結論，’ 爽 ’可以歸類到 ['開心'] 概念中！

舉另一例，
有一詞語 ’ 高興 ’，想知道它是否可歸類於開心概念，於是開始比較：
發現，’ 高興 ’與'快樂'相似度 0.8；與'舒爽'相似度 0.1；與'好玩'相似度 0.1
發現，’ 高興 ’與'快樂'有達到相似度門檻(>0.2)，'舒爽''好玩'則沒有。
也就是’ 高興 ’僅與'開心'概念中的１／３詞語達成相似，換算成小數約 0.33，未達到相似門檻
結論，’ 高興 ’不可以歸類到 ['開心'] 概念中！
即便與該概念中的'快樂'達到 0.8 的相似度，因為與概念整體相似度過低，仍無法被加入該概念中

詳細實現內容可詳閱 GitHub
（https://github.com/spicyscap/AttrClassification/blob/master/Classifaction_M2_Example.ipynb）

C 實作

自定義概念（ 截至 20160117 最新，後續仍有可能變動 ）
（https://github.com/spicyscap/AttrClassification/blob/master/Classifaction_M2_ConceptDic.ipynb）

這次實做上比較了三種不同的 W2V Model，詳細內容如下
（https://github.com/spicyscap/AttrClassification/blob/master/Classifaction_M2_Method_MultiConcept.ipynb）