學號: R05922028 系級: 資工碩二 姓名: 林孟瑤

## 1.PCA of colored faces

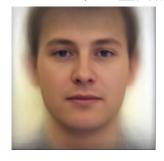
1. (.5%) 請畫出所有臉的平均。

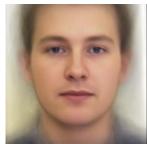


2. (.5%) 請畫出前四個 Eigenfaces,也就是對應到前四大 Eigenvalues 的 Eigenvectors。



3. (.5%) 請從數據集中挑出任意四個圖片,並用前四大 Eigenfaces 進行 reconstruction,並畫出結果。









- 4. (.5%) 請寫出前四大 Eigenfaces 各自所佔的比重, 請用百分比表示並四捨五入 到小數點後一位。
  - 1. 21.6%
  - 2. 10.9%
  - 3. 7.2%
  - 4. 6.1%

## 2. Visualization of Chinese word embedding

1. (.5%) 請說明你用哪一個 word2vec 套件, 並針對你有調整的參數說明那個參數的意義。

#### Genism word2vec:

。 size:向量維度

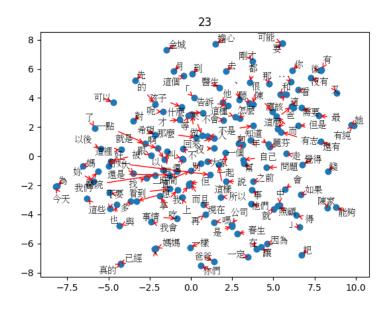
o alpha: learning rate

。 window: 往右看幾個字

。 workers:執行緒數目

。 min\_count: 出現次數超過 min\_count 才會被視為訓練資料

2. (.5%) 請在 Report 上放上你 visualization 的結果。

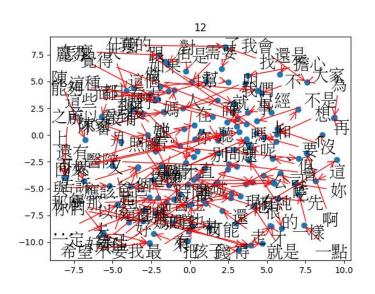


3. (.5%) 請討論你從 visualization 的結果觀察到什麼。

以圖的意義來說明,較有關連的字詞會聚集在同一塊,或是經常一起出現的字 詞也會聚集在一起

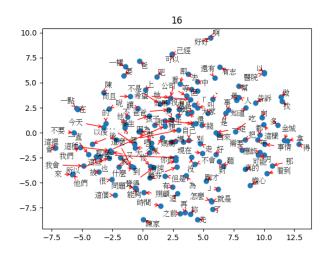
我分別調整了兩個參數去做比較

1. size 也就是維度 這邊降低維度 可以知道當我們維度太高, 則會分的太細 而維度太低的話, 則會分不開來



#### 2. window 調高

在其他參數相同的情況下,將 window 調高,可以讓各點的關聯較多一些



# 3. Image clustering

1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

### 利用 PCA 降維去實作 image clustering 的結果很差:



- 2. (.5%) 預測 visualization.npy 中的 label, 在二維平面上視覺化 label 的分佈。
- 3. (.5%) visualization.npy 中前 5000 個 images 跟後 5000 個 images 來自不同 dataset。請根據這個資訊,在二維平面上視覺化 label 的分佈,接著比較和自己預測的 label 之間有何不同。