

學號：R04522631 系級：機械碩二 姓名：盧玄真

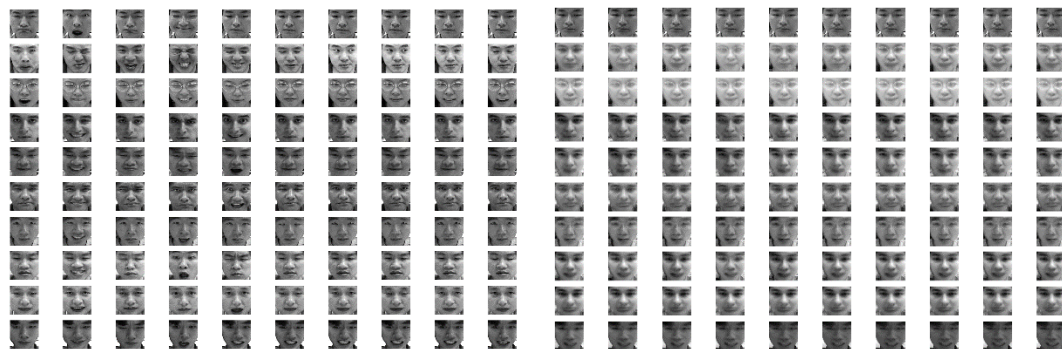
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

答：(左圖平均臉，右圖為 3x3 格狀 eigenfaces, 順序為 左到右再上到下)



1.2. Dataset 中前 10 個人的前 10 張照片的原始圖片和 reconstruct 圖 (用前 5 個 eigenfaces):

答：(左右各為 10x10 格狀的圖，順序一樣是左到右再上到下)



1.3. Dataset 中前 10 個人的前 10 張照片投影到 top k eigenfaces 時就可以達到 $< 1\%$ 的 reconstruction error.

答：(回答 k 是多少)

k 為 11，用 top 11 eigenfaces 就可以使 reconstruction error 達到小於 1%。

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答：

```
word2vec.word2vec("Harry Potter 1-7.txt", "Harry Potter 1-7.bin",
    size=2000, window=10, sample='1e-4', hs=0,
    negative=5, threads= 4, iter_=1000, min_count=10, alpha=0.025,
    debug=2, binary=1, cbow=5, verbose=True)
```

我設定的參數如上：

Size = 2000 代表使用 2000 維的向量來描述一個詞。

Window = 10 表示一個字被相鄰 10 個字影響。因為小說上下文關係重要。

hs = 0 表示不使用 softmax。

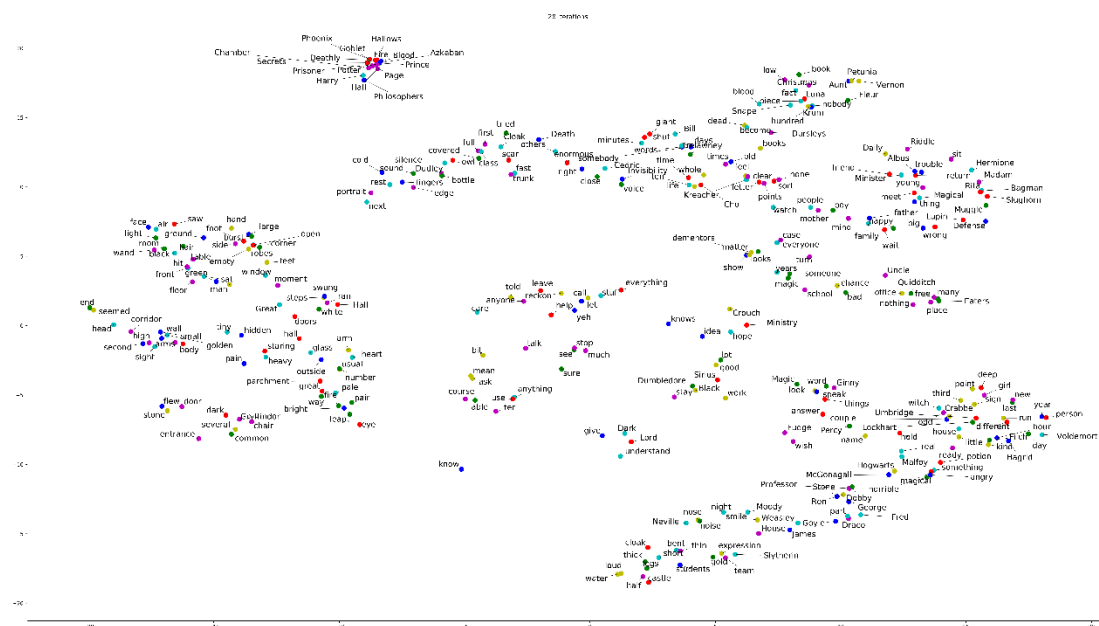
Sample = '1e-4' downsampling 可以降低一些不重要的極高頻詞對模型影響。

Min_count = 10 表示出現 10 次以下的低頻詞不加入訓練。

Alpha = 0.025 表示初始學習率參數。

2.2. 將 word2vec 的結果投影到 2 維的圖:

答：



2.3. 從上題視覺化的圖中觀察到了什麼？

答：

在圖中可以觀察到降維後詞語詞之間點的距離似乎代表著是否經常在同一句話之中出現。最明顯的例子就是，最上面一坨小小的點中的詞有 Harry、Potter、Azkaban、Philosophers...等書的標題，因為書中每一頁最後都會有這些標題，因此經過訓練後並降維距離就很短。而右下角那個群集則是比較多關於佛地魔事件的詞彙例如：Voldemort、Malfoy、Umbridge...等。

3.1. 請詳加解釋你估計原始維度的原理、合理性，這方法的通用性如何？

答：

3.2. 將你的方法做在 hand rotation sequence dataset 上得到什麼結果？合理嗎？
請討論之。

答：