

1. (1 %)請比較有無normalize的差別。並說明如何normalize.

(a)無normalize

public score	0.86992
private score	0.86399

(b)有normalize

public score	0.96738
private score	0.95657

做法:將rating減去平均後再除標準差做為training data，預測的時候再把結果加上平均後乘以標準差，從結果來看normalize並不會比較好

2. (1 %)比較不同的embedding dimension的結果。

dimension	128	256	64	32
public score	0.86992	0.88286	0.86645	0.86836
private score	0.86399	0.87491	0.85903	0.86057

embedding dimension大概落在64~128效果最好，越多不只效果差，而且train要花的時間也比較久

3. (1 %)比較有無bias的結果。

(a)有bias

public score	0.86992
private score	0.86399

(b)無bias

public score	0.91310
private score	0.90793

embedding dimension =128

有bias的效果明顯比較好，原因可能來自每個人平分的基準不盡相同

4. (1 %)請試著將movie的embedding用tsne降維後，將movie category當作label來作圖。

我將movie大概分成五類

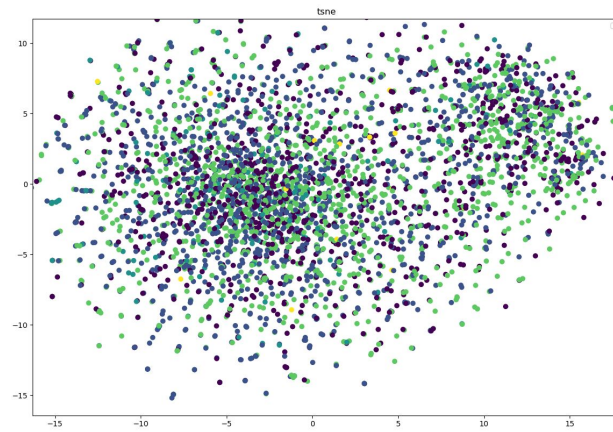
類別1 = ["Action", "Thriller", "Sci-Fi", "Horror", "Crime", "Comedy", "Adventure"]

類別2 = ["Animation", "Children's", "Fantasy"]

類別3 = ["Musical", "Drama"]

類別4 = ["Film-Noir"]

類別5 = 其他



但就圖上來說似乎是沒有很好的分隔開來，幾乎都擠成一團

5. (1 %) 試著使用除了rating以外的feature, 並說明你的作法和結果，結果好壞不會影響評分。