

ML hw6 report

學號：B05611033 系級：生機二 姓名：杜杰翰

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

我的model是將user, movie做embedding到128維之後concatenate起來，然後丟入DNN得到一個一維輸出，之後在將其與user, movie的bias加起來。

在這種情況下，有無normalize在kaggle上的結果如下：

	private	public
有 normalize	0.86198	0.86912
無 normalize	0.86381	0.87011

我將ratings減去平均之後除以標準差，在做完predict之後再乘上原本的標準差並加上平均。

2. (1 %)比較不同的embedding dimension的結果。

我使用第一題中有normalize的model進行了64, 128, 256，三種不同維度的試驗，在kaggle上的結果如下：

	private	public
64維	0.86570	0.87025
128維	0.86198	0.86912
256維	0.86719	0.87209

結果發現較大的維度不一定能得到較好的結果，有可能是因為在過大的維度中，有些維度對結果並無關聯，但還是會一定程度影響結果的緣故。

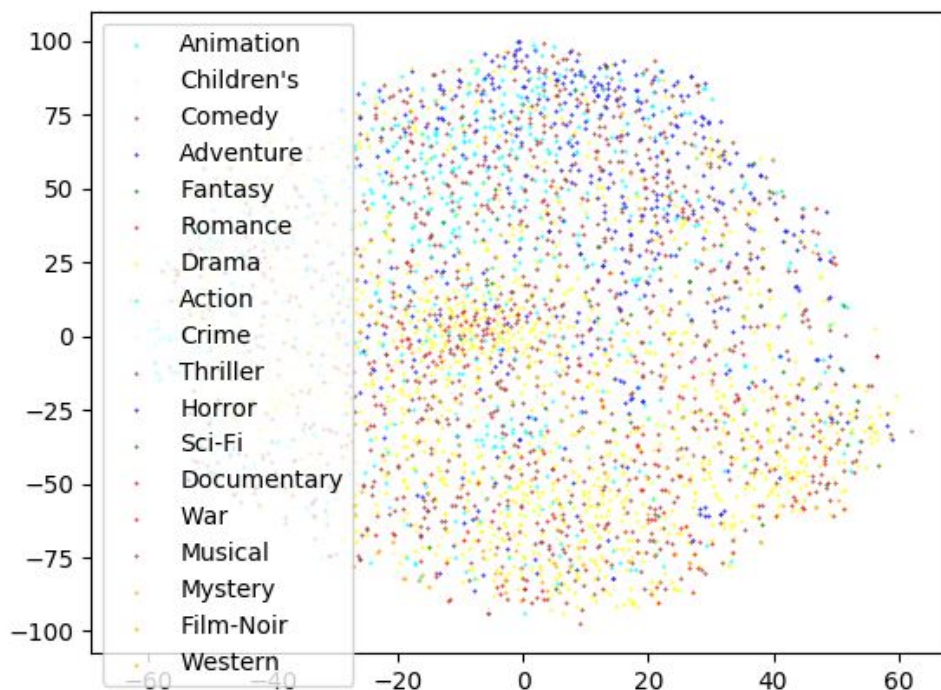
3. (1 %)比較有無bias的結果。

我使用第一題中有normalize的model進行了有無bias的試驗，在kaggle上的結果如下：

	private	public
有 bias	0.86198	0.86912
無 bias	0.87786	0.88395

user與movie的bias代表著他們本身觀看電影的欲望與對觀眾的吸引力。跟預測相同，再沒加bias的時候正確率會產生明顯下降。

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



我用embedding到32維後用tsne降至2維，並取各movie的第一個category來當label。

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。

我將users.csv的Gender, Age, Occupation進行onehot，將movies.csv的年份與類型進行onehot，然後分別將他們經過一linear layer到128維之後與embedding到128維的users, movies進行concatenate，然後經過DNN之後輸出至一維。其結果在kaggle表現如下：

public : 0.86932

private : 0.86440

結果看來與只使用rating差不多，可見用戶及電影的那些feature可以經由training得到差不多的結果。