

Report

學號：B05902002 系級：資工二 姓名：李栢淵

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

無 normalize：

try7.csv 3 days ago by Bai-Yuan Lee add submission details	0.84522	0.85341	<input checked="" type="checkbox"/>
---	---------	---------	-------------------------------------

有 normalize：

try8.csv 3 days ago by Bai-Yuan Lee add submission details	0.84444	0.85301	<input type="checkbox"/>
---	---------	---------	--------------------------

有 normalize 會比較好一點，可能是 predict 的範圍變比較小，能夠 train 的比較細。

方法：

將原本所有的 rating 計算其平均及標準差，進行標準化（減去平均，再除以標準差），再進行 training 及 predict，最後再乘上標準差、加回平均。

（註：這裡使用的方法是加上rating之外的feature的方法，細節參考第五題）

(1 %)比較不同的 embedding dimension 的結果。

embedding 到 18 維：

try2.csv 10 days ago by Bai-Yuan Lee add submission details	0.86741	0.87382	<input type="checkbox"/>
--	---------	---------	--------------------------

embedding 到 30 維：

try3.csv 10 days ago by Bai-Yuan Lee add submission details	0.85752	0.86540	<input type="checkbox"/>
--	---------	---------	--------------------------

embedding 到較高維度表現得較好，因為model的複雜度提升了，但再高就很容易 overfit。

（註：這裡使用的方法非MF，而是DNN）

(1 %)比較有無 bias 的結果。

無 bias：

[try1.csv](#)

11 days ago by [Bai-Yuan Lee](#)

[add submission details](#)

0.87518

0.88471



有 bias：

[try4.csv](#)

5 days ago by [Bai-Yuan Lee](#)

[add submission details](#)

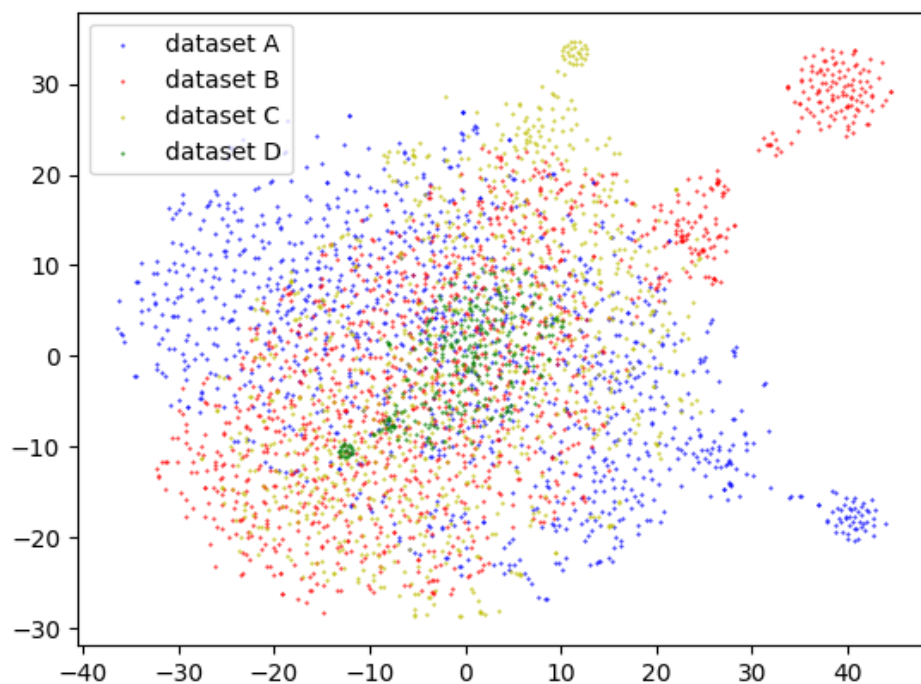
0.86599

0.87490



有 bias 的結果會好滿多的，因為多考慮了每一個 user 跟 movie 各自的 feature。

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



Dataset A：Animation，Children's，Comedy。（開心類，藍色）

Dataset B：Adventure，Fantasy，Romance，Drama。（奇幻類，紅色）

Dataset C：Action，Crime，Thriller，Horror。（刺激可怕類，黃色）

Dataset D：Sci-Fi，Documentary，War，Musical，Mystery，Film-Noir，Western。（其他類，綠色）

(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。

沒有其他 feature 的純 MF：

try4.csv 5 days ago by Bai-Yuan Lee add submission details	0.86599	0.87490	<input type="checkbox"/>
--	---------	---------	--------------------------

加了其他feature的DNN：

try5.csv 4 days ago by Bai-Yuan Lee add submission details	0.84631	0.85437	<input type="checkbox"/>
--	---------	---------	--------------------------

把 user 跟 movie 的 feature 變成 one-hot encoding，先接一層 Dense 到跟 latent vector 一樣的長度，把他跟之前 MF embedding 出來的東西 concat 起來，直接接 DNN，結果變好滿多的，我就有用此方法繼續往更高的 embedding dimension 嘗試。