

r06922115

鄭皓謙

HW2 report

Task 1

這份報告是task1截止後，跟同學討論發現自己題目理解錯誤之後，才重新做的。我沒有注意到題目是要預測使用者「沒吃過」的食物，所以我用與task1相同的做法，沒有加入同學報告的做法，去掉那些吃過的時候以後，丟late submission的結果，希望多少能補救成績，

Validation

資料中最長的序列為不連續160不同的日期的飲食，最短為25個日期，我將每個user的最後7天當作validation去試不同的model，計算accuracy。

ALS, BPR

這是這次作業的首先嘗試的做法，先將每個紀錄做成cof matrix，試過implicit 0/1的ranking matrix，也試過存放吃的次數，以ALS來說 implicit效果較好，而後者則是BPR比較好。(BayesianPersonalizedRanking)

還有一些ALS的變形，Annoy ALS, Faiss ALS，但是這些在validation的時候效果就不好了，最後沒有加入。

找到在validation上最好的模型參數：ALS factor:32 iteration : 210 accuracy 0.5808

在kaggle上的late submission:

Private Score	Public Score
0.04302	0.04254

SVD

嘗試使用SVD分解implicit Ranking matrix，再找出最常吃的食物，與最常吃食物的cosine similarity。效果還不錯0.03845/0.04030

嘗試與心得

起初很多做法分數都很低，像是SDAE，跟同學討論說他用SVD就過baseline我也覺得很驚訝，剛開始還懷疑是不是negative sampling有錯誤，後來才發現是題目理解錯誤，希望late submission可以補救。