twu38-HW4

Tianqi Wu

10/31/2019

Problem 1

Display network targets by certain types of people of certain contexts. First, we can obtain user's past search queries. Then, we can use Simrank to find similar queries to rewrite those queries. Finally, we can use the rewrote queries to display relevant ads.

Problem 2

λ_2 and λ_3 are regularization parameters which can shrunk the average. They also serve as smoothing factors that reduce bias when the number of users or items is small. They are often determined by cross validation.

Problem 3

One latent factor is equivalent to a most-popular recommender. As number of latent factors increase, personalization and model quality would be improved until it overfits. We can determine the number of latent factors by line search or grid search that gives optimal training and test accuracy.

Problem 4

The missing values pattern is not random since people tend to rate items that impress them. With binary dense view, the model characterises which items users rated, rather than how they rated. User factors are indirectly defined by item factors and we can identify the co-association relationships between users and items. The model would be better at predicting users' preference and more robust when calculating bias. By integrating the rating model and dense binary view, the model is able to produce a 7% improvement.

Problem 5

Pearson correlation

user	item	score
15	7	3.3
15	177	2.6
19	921	2.5
20	1089	4.6
21	1393	2.5

Content similarity

user	item	score
15	7	2.3
15	177	2.2
19	921	3.8
20	1089	3.7
21	1393	3.3