Movie2 Summary

Jinfeng Lin

February 8, 2015

1 URL

First of all is the URLs
code climate : `https://codeclimate.com/repos/54d7dea0e30ba06d77000cc5/feed`
Git: `https://github.com/jinfenglin/movie2/tree/master`

2 Prediction Algorithm

The Prediction algorithm here we use is mostly a user-based algorithm. The algorithm will first get the similarity vector for user U with other users. In this step, I made a slight modification to improve the speed, that I only get the similarity between user U and other viewer who had watched movie M, because if users have not seen the Movie M, they could not provide useful information to predict. Then I chose k users who has the highest similarity with user U and get their average score on movie M and use this result as the final predication for user U on movie M. The benefit for this algorithm is that it is easy to implement, which is based on the similarity method that we have implemented in movie1. The drawback is that it is not quite fast and have low accuracy.

3 Analysis

The final statistic report is as follow:

Mean	RMS	STDDEV
0.81	3.5	2.7

Notice that the Mean here indicate the absolute error which is equal to —real - predict—, so the higher estimation in one instance will not balance another estimation which is too low. The average error is about 0.8 which is quite high, increasing the similarity list length will not help to control it. So may be some other algorithm are needed if we want higher accuracy. I suppose one problem here is the similarity method, it could not really find the most similarity ones which lead to low accuracy in prediction.

4 Benchmarking

The time of making a single prediction is about 0.02 s in average. If the training set is expended the prediction methods definitely will takes more time because it need to go through more instance to make similarity calculation.