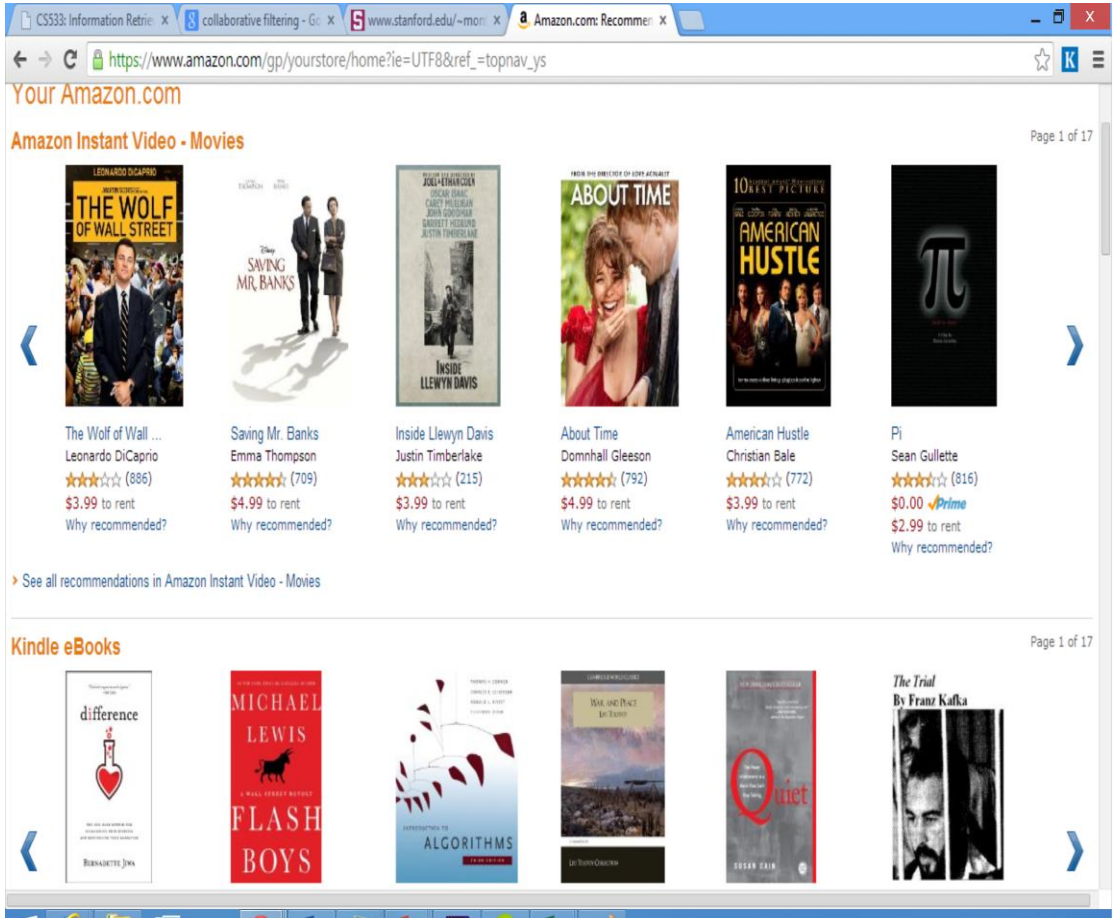
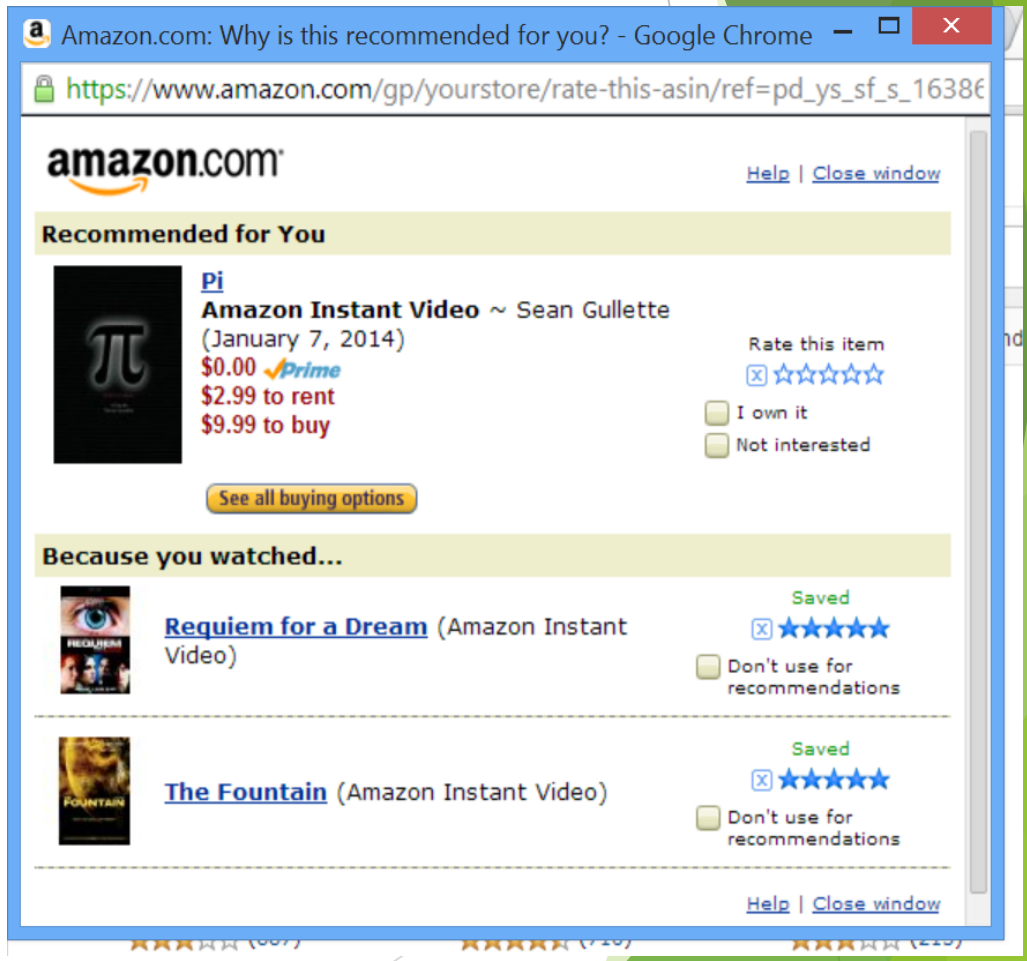
Movie Recommendation System

Neslihan Bulut

Recommendation Systems

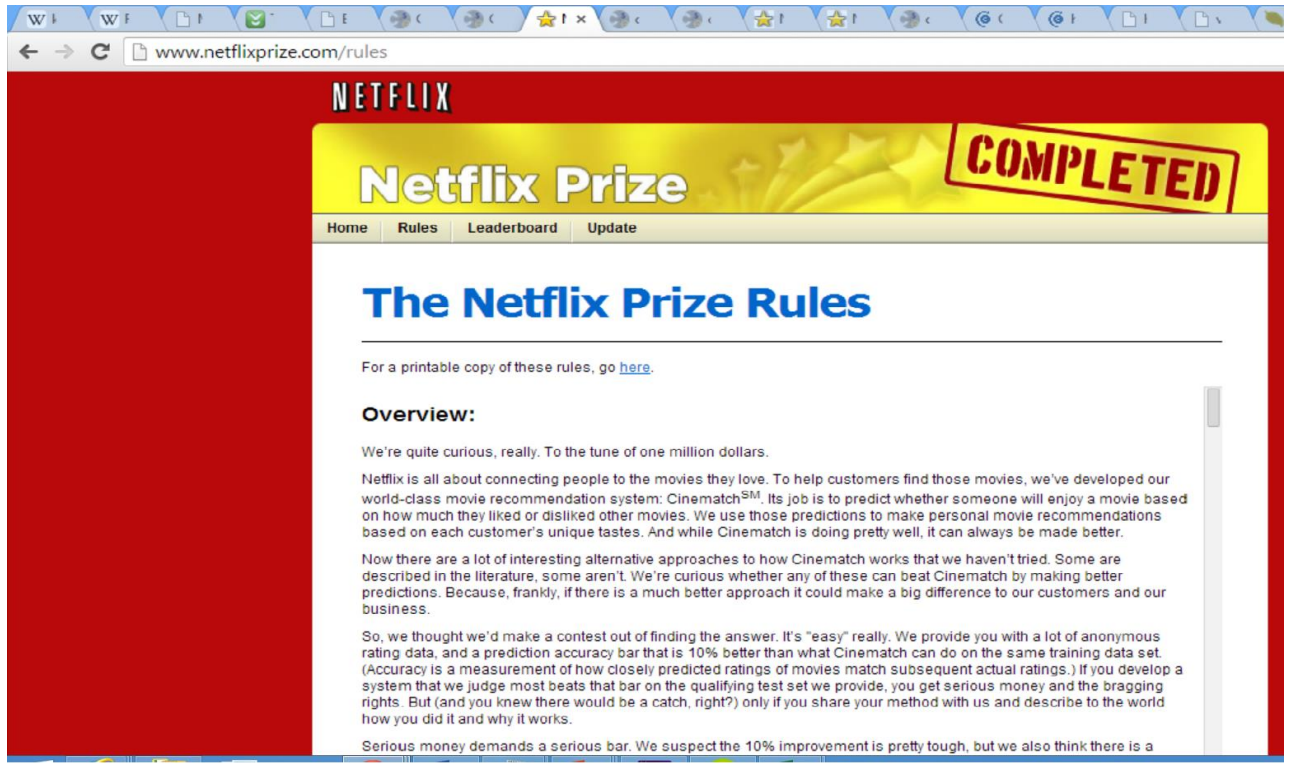




Collaborative Filtering

- Predictions on how well a user will like an item
- Has a training data of items, users and user votes
- Neighborhood methods are used ; find similar users and make predictions based on their votes

Netflix Data



Netflix Data

- 3M votes for the training data
- 100K votes for testing the accuracy of the algorithm

Algorithms: Collaborative Filtering (Breese et al.)

- $v_{i,j}$= vote of user *i* on item *j*
- I_i = items for which user *i* has voted
- Mean vote for *i* is

$$\overline{v}_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{i,j}$$

- Predicted vote for "active user" *a* is weighted sum

$$p_{a,j} = \overline{v}_a + \kappa \sum_{i=1}^{n} w(a,i)(v_{i,j} - \overline{v}_i)$$

normalizer

weights of *n* similar users

Algorithms: Collaborative Filtering (Breese et al.) – Cont.

- Pearson correlation coefficient (Resnick '94, Grouplens):

$$w(a,i) = \frac{\sum_j (v_{a,j} - \overline{v}_a)(v_{i,j} - \overline{v}_i)}{\sqrt{\sum_j (v_{a,j} - \overline{v}_a)^2 \sum_j (v_{i,j} - \overline{v}_i)^2}}$$

- Most similar users are selected to predict scores for unrated movies
  - *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*, Breese, Heckerman, Kadie

Evaluation

- Test and training sets from the Netflix prize
- Compare the predictions from the algorithm with the actual vote data from the test set

Thanks for listening!

- Any questions?