Collaborative Filtering

Group 4: Lin Han, Vassily Carantino, Xinyao Guo, Qihang Li, Qian Shi

Layout:

1) Purpose of the study and approaches used

2) How to evaluate the performance of our different algorithms ?

3) Steps we took in our experimental approach

4) Finding the best Memory-Base Model

5) Finding the best Cluster Model

6) Results & Analysis: What is the best model overall

1) Purpose of the study and approaches we used

Goal of our Study

→ We want to compare different options to perform collaborative filtering on two distinct database.

→ We need to compare the following things:

- Memory-based Algorithm(based on neighbors)
  - ✓ Different combinations of components
- Model- based Algorithm
  - ✓ Choose the best meta-parameters

Experimental Approach

Here are the different steps of our experimental approach:

1. Find the best Memory-based Algorithm(based on neighbors)
   - ✓ Do predictions based on different combination of components
   - ✓ Evaluate the test data to find best combinations of components
2. Find the best Model-based Algorithm
   - ✓ Do predictions based on different parameters
   - ✓ Evaluate the test data to find the best parameters
3. Compare the two and come up with a recommended algorithm for a specific usage.

2) How to evaluate the performance of our different algorithms ?

1 Web Dataset

Web Dataset: Ranked Scoring: we default alpha=5, d=0.5

$$R_a = \sum_j \frac{\max(v_{a,j} - d, 0)}{2^{(j-1)/(\alpha-1)}}$$

$$R = 100 \frac{\sum_a R_a}{\sum_a R_a^{max}}$$

2 Movie Dataset

1. MAE = ABS(Estimation – real score)
2. Multi-class ROC:

We extend the definition of ROC defined by Hand and Till, to more than 2 classes by averaging pairwise comparisons. This measure reduces to the standard form in the two class case.

3) Steps we took in our experimental approach

1

Memory-based Algorithm

dis-similarity to similarity.

Similarity Weight:

- Pearson Correlation:
- Entropy: 1,2
- Mean-Square-difference: 1,2
- SimRank: 1

Significance Weighting: 1,2

Selecting Neighbors:

- Weight Threshold: 1,2
- Best-n-estimator: 1,2
- Combined: 1,2

Rating Normalization:

- Deviation for Mean:1,2

Evaluation: Ranked Scoring for d1, MAE for d2

2

Model-based Algorithm

EM Algorithm(Log-likelihood Function)

Score Estimation

Evaluation: MAE, ROC

4) Finding the best Memory-based Model

1 MAE for movie Data of Memory-based Algorithm

Similarity Weighting	Significance Weighting	Selecting Neighbors	MAE
Pearson	F	Best-N(10)	3.03
Pearson	F	Best-N(15)	2.97
Pearson	T	Weight Threshold(0.4)	2.64
Pearson	T	Weight Threshold(0.5)	2.43
Pearson	T	Combined	2.43
MSE	F	Best-N(10)	3.28
MSE	F	Best-N(15)	3.24
MSE	F	Combined	3.18
MSE	T	Best-N(10)	2.66
MSE	T	Best-N(15)	2.62
MSE	T	Weight Threshold(0.85	2.42
MSE	T	Combined	2.57
Entropy	F	Best-N(10)	3.41
Entropy	F	Weight Threshold(0.3)	2.9
Entropy	F	Combined	3.31
Entropy	T	Best-N(10)	2.83
Entropy	T	Best-N(15)	2.64
Entropy	T	Weight Threshold(0.3)	2.6
Entropy	T	Combined	2.7

Analysis

- For the Movie Dataset, Significance Weighting seems to be a need to reduce MAE in every combination. In terms of Selecting Neighbors, Weight Threshold outperforms Best-N and Combined Method. For best-N, n=15 outperform n=10.
- Specifically, MSE+Significance Weighting+Weigth Threshold(0.85) performs the best. Pearson+Significance Weighting+Weight Threshold(0.5) and Pearson+Significance Weighting+Combined also perform well.

4) Finding the best Memory-based Model

2 Ranked Scoring for Web Dataset

Similarity Weighting	Significance Weighting	Selecting Neighbors	Ranked Score
Pearson	F	Best-N(10)	53.07335993
Pearson	F	Best-N(15)	53.88041758
Pearson	F	Combined	54.82918581
Pearson	F	Weight Threshold(0.4)	61.05223056
Pearson	F	Weight Threshold(0.5)	58.98323061
Pearson	F	Weight Threshold(0.6)	52.6233943
Pearson	T	Best-N(10)	53.07335993
Pearson	T	Best-N(15)	53.88041758
Pearson	T	Weight Threshold(0.5)	58.98323061
Pearson	T	Weight Threshold(0.6)	52.6232261
Pearson	T	Combined	55.83386888
MSE	F	Best-N(10)	57.43879672
MSE	F	Best-N(15)	62.59660348
MSE	F	Weight Threshold(0.08)	69.10069659
MSE	F	Weight Threshold(0.1)	59.65513169
MSE	F	Combined	69.13923115
MSE	T	Best-N(10)	57.29309379
MSE	T	Best-N(15)	61.86672917
MSE	T	Weight Threshold(0.1)	58.38028125
MSE	T	Combined	68.70204497
Entropy	F	Best-N(10)	51.77175435
Entropy	F	Best-N(15)	53.78706305
Entropy	F	Weight Threshold(0.8)	61.44838698
Entropy	F	Combined	55.57004837
Entropy	T	Best-N(10)	51.77175435
Entropy	T	Best-N(15)	53.78706305
Entropy	T	Weight Threshold(0.7)	74.0076065
Entropy	T	Combined	55.94488558

Analysis

- For the Microsoft Dataset, in terms of Similarity Weighting, MSE and Entropy outperforms Pearson Correlation in general. In terms of Selecting Neighbors, Weight Threshold seems to be the best in most combinations.
- Specifically, Entropy+Significance Weighting+Weight Threshold(0.7) performs the best. MSE+Combined and MSE+Weight Threshold(0.7) also perform well.

5) Finding the best Cluster Model

- Before comparing the Cluster model with the memory-base algorithms respective performance on the Movie Data set we need to train the Cluster model

1 Idea of the Cluster model

- **idea:** there are certain groups or types of users capturing a common set of preferences and taste.
  - Given a unobserved class variable C, the votes of users for a particular movie are iid
  - We can use the EM algorithm to learn the parameters:

$$\mu_c := P(\Delta_i = c), \quad \text{for} \quad c = 1, \ldots, C;$$

$$\gamma_{c,j}^{(k)} := P(V_j^{(i)} = k | \Delta_i = c), \quad \text{for} \quad \forall c, j, k.$$

2 Training the Cluster Model:

- **Method to find the best number of cluster C**
  - We divide the training set into a sub-training set and a validation set. In the validation set, 70% of the each user’s votes are known, the remaining 30% are predicted.
  - For each value of C we run the EM algorithm on the sub-training set, and we evaluate the performance metrics on the validation set.
  - We obtain the following results, the best parameter is C=7

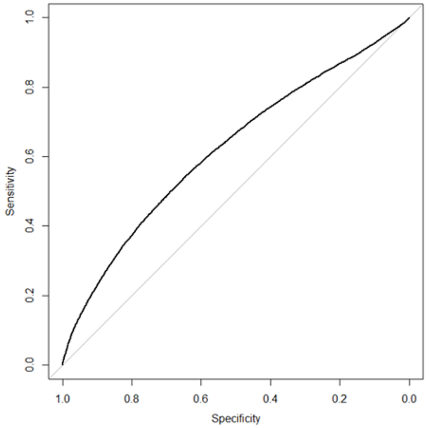
C	3	5	7	9
MAE	1.01345	1.00694	1.00455	1.021888
ROC	0.2474	0.2467	0.2445	0.2535

- **Testing performance:**
  - We ultimately test the performance of the model on the test set with C=7

6) Results & Analysis: Comparing all the models

1 Cluster Models Performance

- We choose C=7, train on the whole training set, and evaluate on the testing set.
- **The results are: MAE=0.99195, ROC=0.2397**



2 Comparing the performance of all the methods on the Movie Data

Similarity Weighting	Significance Weigh	Selecting Neighbo	MAE
Pearson	F	Best-N(10)	3,03
Pearson	F	Best-N(15)	2,97
Pearson	T	Weight Threshold(0.4	2,64
Pearson	T	Weight Threshold(0.5	2,43
Pearson	T	Combined	2,43
MSE	F	Best-N(10)	3,28
MSE	F	Best-N(15)	3,24
MSE	F	Combined	3,18
MSE	T	Best-N(10)	2,66
MSE	T	Best-N(15)	2,62
MSE	T	Weight Threshold(0.8	2,42
MSE	T	Combined	2,57
Entropy	F	Best-N(10)	3,41
Entropy	F	Weight Threshold(0.3	2,9
Entropy	F	Combined	3,31
Entropy	T	Best-N(10)	2,83
Entropy	T	Best-N(15)	2,64
Entropy	T	Weight Threshold(0.3	2,6
Entropy	T	Combined	2,7
Cluster-Model using C=7			0,99

Recommendations

The best collaborative filtering algorithm on this data set is our **Cluster Model with 7 latent classes.**

Let’s note that might not be the case on the Web Data Set, for which our better **tested** algorithm is the memory-based Entropy+Significance Weighting+Weight Threshold(0.7). We will need to run our cluster model on this data base to conclude.