

## AUTOMATED COLLABORATIVE FILTERING

Reduce information load

Complements Content-based information filtering systems

Collaborative filtering utilizes human judgments(ratings)

Key advancements over content based filtering:

- Support for contents not easily analyzed by automated processes
- Filter items based on quality and taste
- Ability to provide serendipitous recommendations

## NEIGHBORHOOD BASED METHODS

Neighborhood based methods has 3 basic steps;

- Weight all users with respect to similarity with the active user
- Select a subset of users to use as a set of predictors
- Normalize ratings and compute a prediction from a weighted combination of selected neighbors' ratings

One of the first introduced neighborhood based automated collaboration filtering algorithm;

- Pearson correlation to weight user similarity;

$$w_{a,u} = \frac{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sigma_a - \sigma_u}$$

- Use all available correlated neighbors
- Compute a final prediction: weighted average of deviations from the neighbor's mean:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

## EXPERIMENTAL TECHNIQUE

Compare results of different neighborhood based prediction algorithms

Data: anonymous reviews from the MovieLens movie recommendation site

- 122,176 ratings from 1173 users, every user having at least 20 ratings.
- %10 users selected to be test users, ratings for 5 items were withheld
- For each item getting predicted highest ranking neighbors that have rated the item are used for the prediction

Quality of a prediction algorithm;

- Coverage: Usually high
- Accuracy: Statistical accuracy or decision support accuracy

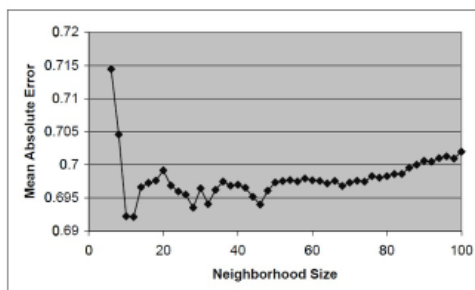
## WEIGHTING POSSIBLE NEIGHBORS

Similarity weighting

- Pearson Correlation Coefficient: Measures the degree to which a linear relationship exists between two variables.
- Spearman Rank Correlation Coefficient: Similar to Pearson but doesn't rely on model assumptions, and performs similarly as well.
- Vector Similarity: performs well for information retrieval but not as good for collaborative filtering
- Entropy: Not as good as pearson correlation
- Mean-squared difference: Not as good as pearson correlation

## SELECTING NEIGHBORHOODS

Select a subset



## SELECTING NEIGHBORHOODS

### Best-n-neighbors

- Pick best n correlates
- Large n => too much noise
- Small n => poor predictions

## PRODUCING A PREDICTION

Combine ratings of neighbors

Deviation-from-mean approach