An Algorithmic Framework for Performing Collaborative Filtering

Automated Collaborative filtering works by human judgements such as ratings for item. Then the users are matched who share the same interest in different areas such as their opinions on different items, a weight is assigned to each user which in accordance with the active user as how similar that user is to active user, more weight means more similar and hence is more suitable predictor and less weight means less similar and hence less similar and hence less suitable for predictor.

Collaborative filtering follows a neighbourhood-based method:

1. Weight all users with respect to similarity with the active user

* Similarity weighting:

Automated Collaborative filtering works by human judgements such as ratings for item. Then the users are matched who share the same interest in different areas such as their opinions on different items, a weight is assigned to each user which in accordance with the active user as how similar that user is to active user, more weight means more similar and hence is more suitable predictor and less weight means less similar and hence less similar and hence less suitable for predictor.

* Significance weighting:

It depends on how many common items have they rated because if the number of items both have rated is less than it affects the quality of results, so for this reason more weight is given to the relationships where they have more common items between them.

* Variance Weighting:

Not proved to be helpful maybe because during the hypotheses they didn’t account for some important factors. But the main idea was to select items that actually affect the decision to separate taste between users rather than going with the most common items.

Note: Spearman correlation performed better than Pearson correlation because it is not dependent on model assumption.

1. Select a subset of uses to use as a set of predictors

* Two ways of doing that:

1. Keeping a threshold value and whichever crosses than threshold value will contribute in the prediction.
2. Picking n -best Users

Note: N – best proved to be much more effective

1. Normalize ratings and compute a prediction from a weighted combination of selected neighbour items.

* From all the approaches available deviation-from-mean proved to increase accuracy significantly over weighted average or z-score averaging which provided no significant improvement.

Two types of data can be used in Collaborative Filtering

Explicit : Are generally a single numeric summary like the rating for each item

Implicit: Drived from already present data sources like purchase records, web logs, time spent reading