Report Recommender Systems COMP40320
Item-based ACF Recommendation
Remi Pichon 15208089

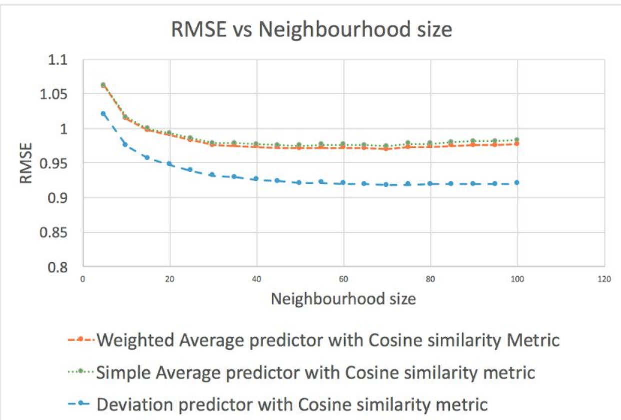
This item-base ACF recommendation is based on movies rated by users. The system attempts to predict the rating of users for movies according to a training dataset.
RMSE and coverage are used to show prediction quality. RMSE means Root Mean Squared Error and is used to quantify the error on the rating precision. Coverage refers to the percentage of how many predictions have been computed.

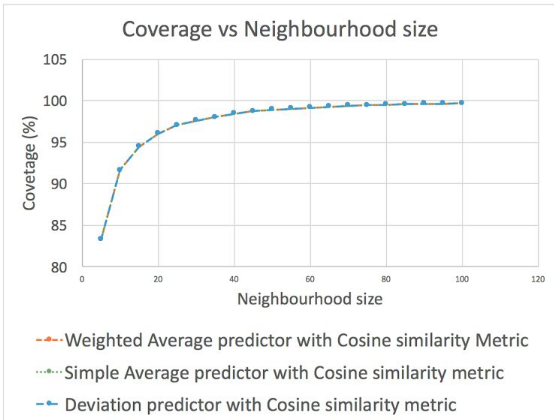
Effect of neighbourhood size on predictions

The experiment

Using the nearest neighbourhood approach and cosine similarity metric this experiment shows three predictor results for range of neighbourhood sizes. Neighbourhood item size varies from 5 to 100. Simple average, weighted average and deviation from item-mean are the predictors used in this experiment.

The results





On the RMSE graph the upper line (the worst) is the Simple Average. Just a little better is the Weighted Average and the lower line (the best) is the Deviation predictor.
On the Coverage graph, the three predictor give the exact same coverage.
The Deviation predictor is an average 5% better RMSE than the Weighted predictor. Weighted average is less than 0.4% better than Simple average.

The analysis

First of all, the coverages are strictly equals because the same similarity (cosine) is used. Indeed, the limiting factor for the coverage is the formation of the neighborhood which depends on the similarity.

A neighbourhood size of 50 is the best compromise. The RMSE is 0.9210 and the coverage 98.907%. With a neighbourhood twice the size (100), the RMSE is 0.9202 (8% better than for 50) and the coverage is 0.75 points better. With a neighbourhood two times bigger the memory footprint is two times bigger as well as computing time.

Using the Weighted Average predictor, the standard deviation for every item similarity weight of one user has been computed. These values show how close in value the item similarity weights are.
The average of all these standard deviation measurements (one per user) is 0.036370. This result means that when users co-rated movies, they co-rated movies who were already very similar.
The standard deviation of all these standard deviation is 0.0144, showing that this analysis concerns every users.

Neither Weighed nor Simple Average rely on normalisation as Deviation from Mean Average do. By taking into account the average rating of the target item and the average rating of every item in its neighbourhood, Deviation from Mean is more robust against items always highly or lowly rated. Thus, Deviation predictor outputs better predictions.

$$p_{a,j} = \frac{\sum_{i=1}^{n} w_{a,i}\; r_{i,j}}{\sum_{i=1}^{n} |w_{a,i}|}$$

Weighted Average predictor

$$p_{a,j} = \bar{r}_a + \frac{\sum_{i=1}^{n} w_{a,i}\; (r_{i,j} - \bar{r}_i)}{\sum_{i=1}^{n} |w_{a,i}|}$$

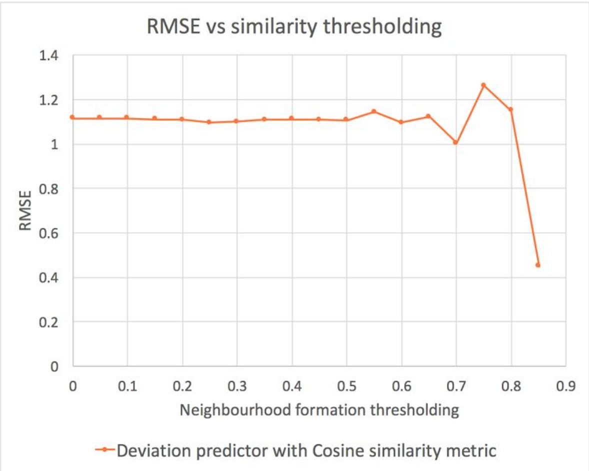
Deviation from Mean predictor

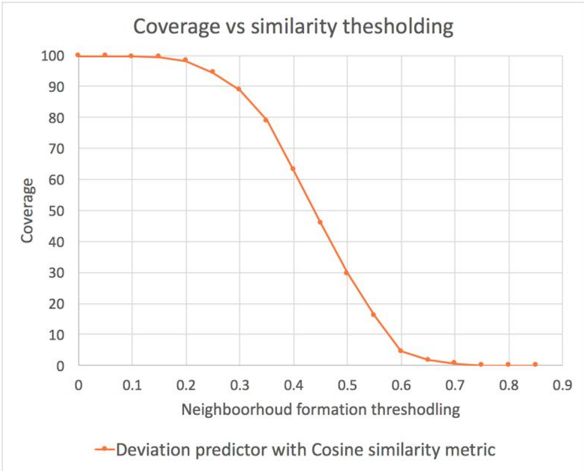
Effect of neighbourhood threshold on predictions

The experiment

Using the deviation from item-mean predictor and cosine similarity metric, this experiment shows the effect of the similarity threshold when building neighbourhood. Threshold varies from 0 to 0.8.

The result





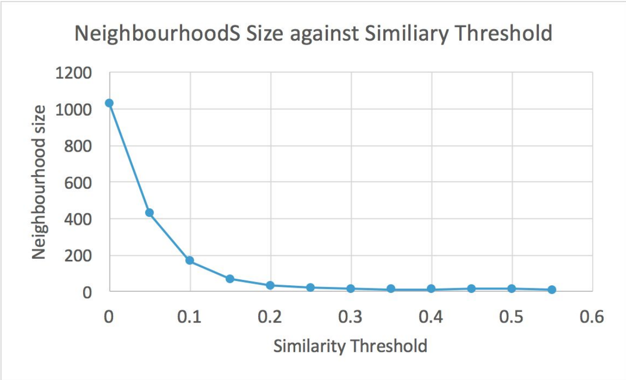
RMSE is constant to 1.1 for a threshold smaller than 0.5. RMSE above a similarity thresholding of 0.7 gives unexpected results.

Coverage starts decreasing for thresholds greater than 0.2. Coverage over a threshold of 0.4 is less than 50%.

The analysis

For a similarity threshold of 0.8, the RMSE is excellent (0.45) but the coverage is close to 0%. A threshold of 0.8 gives an average neighbourhood size of 12 while a threshold of 0.2 give an average size of 148 items.
With a high threshold, the size of the neighborhood is too small to be relevant and leads to an artificial decrease of RMSE, as the predictions only concerns a few movies. The system predicts good rating but only for specific items.

A high threshold is needed to have a smaller neighbourhood to spare computing resources. Between 0 and 0.5 the RMSE is constant (less than 0.006 standard deviation) but the coverage goes under 30% for a threshold at 0.5.



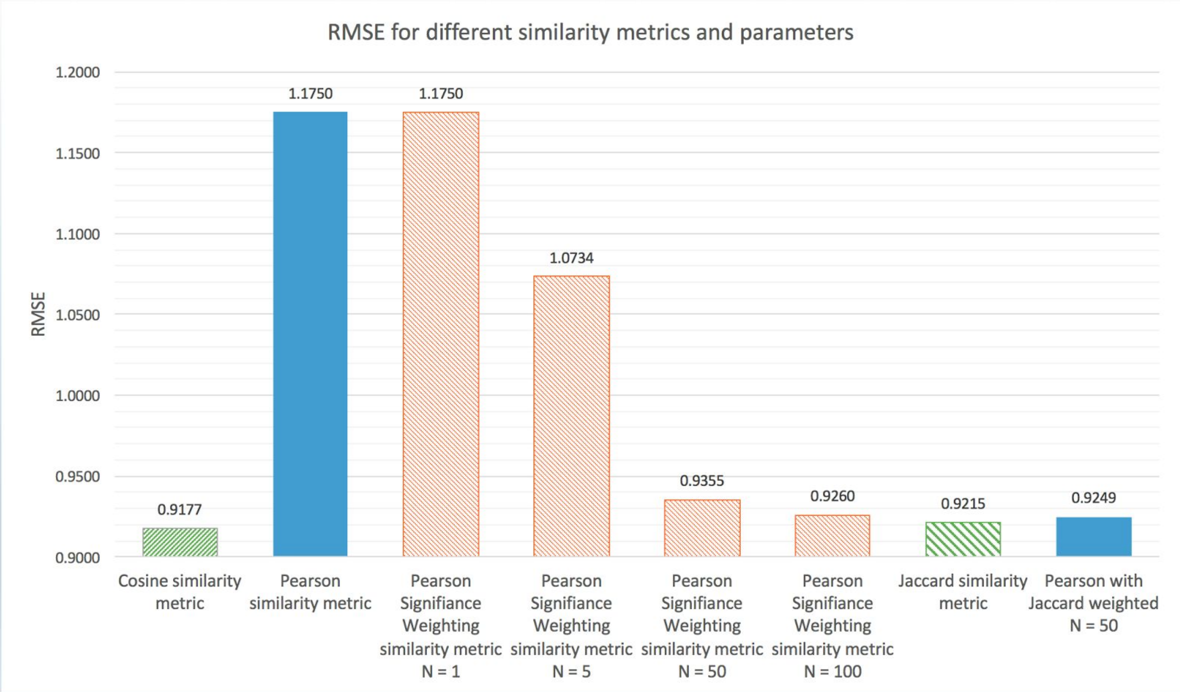
A compromise between neighbourhood size and coverage should be found to spare resources. A threshold of 0.2 has the best result : RMSE is 1.1081, coverage is 98.28% and neighbourhood size is an average 33 items.

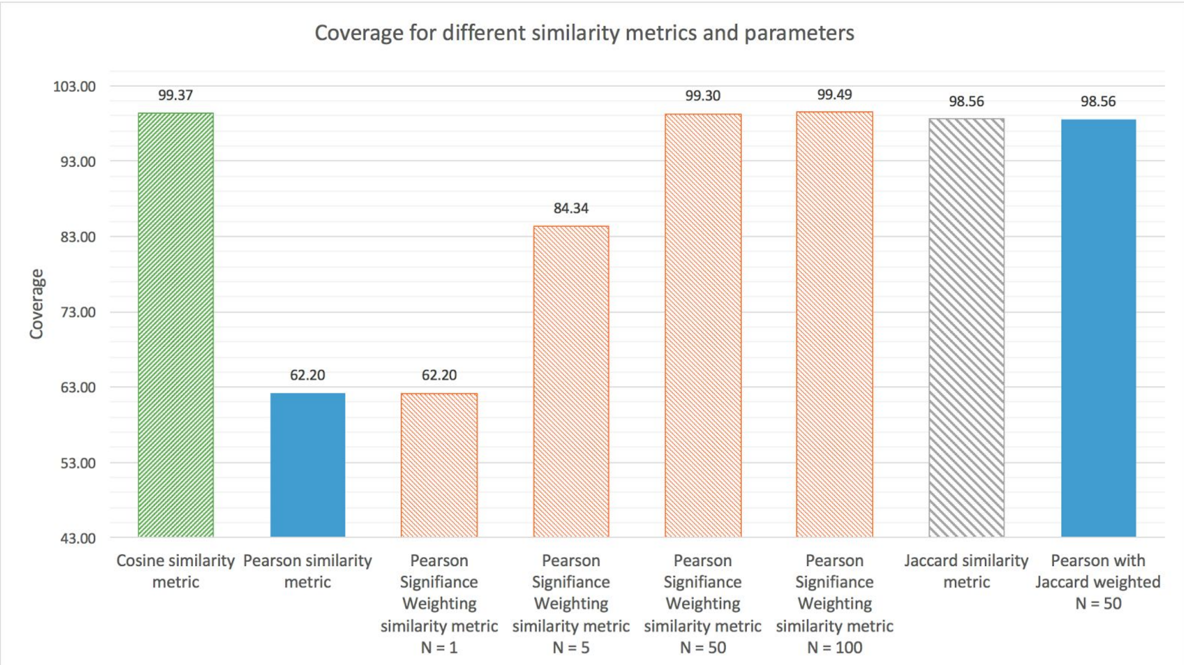
Effect of similarity metric on predictions

The experiment

Using the deviation from item-mean predictor and nearest neighbourhood formation with a size of 70, this experiment compares the following metrics : Pearson correlation, Pearson correlation with significance weighting, Cosine and Jaccard.

The result





Pearson with significant weight get lower RMSE and higher coverage with an increase of the parameter N. The significantly better results are bounded by N = 50.

Cosine and Jaccard both have accurate predictions along with excellent coverage.
Pearson with N = 100 has the best coverage but Cosine and Jaccard give more precise ratings.

The analysis

Pearson metric (without significant weight) produces less accurate similarities than Pearson with significant weighting. As parameter N increases, so does the precision in the neighbourhoods built, which leads to a better accuracy when computing rating.
N is responsible for lowering the impact of items that are not similar enough. When N increases, the more similar co-rated movies are taken into account and the others influence the result less With N = 50, only the most relevant items are used to predict the ratings.

Both Cosine and Jaccard rely on more than only summations over co-rated items as Pearson does. They take a look at every user's rating. Indeed, both formulas use the intersection (co-rated movies) on the numerator and the whole ratings on the denominator. Insofar as they are using the same sets of data they both predict with the same precision which is better than the precision obtained using that Pearson metric.

$$w_{a,i} = \frac{\sum_{j\in I_a\cap I_i} r_{a,j}\ r_{i,j}}{\sqrt{\sum_{k\in I_a} r^2{}_{a,k}}\ \sqrt{\sum_{k\in Ii} r^2{}_{i,k}}} \qquad J_{a,i} = \frac{|a \cap i|}{|a \cup i|} \qquad w_{a,i} = \frac{\sum_{j\in I_a\cap I_i}(r_{a,j}-\bar{r}_a)(r_{i,j}-\bar{r}_i)}{\sqrt{\sum_{j\in I_a\cap I_i}(r_{a,j}-\bar{r}_a)^2\ \sum_{j\in I_a\cap I_i}(r_{i,j}-\bar{r}_i)^2}}$$

Pearson Jaccard Cosine

Cosine depends on the actual ratings but Jaccard only uses the set size to perform similarity. As both Cosine and Jaccard are the most accurate, a conclusion can be drawn : the actual values of the movie ratings do not play an important role when making predictions. Instead, it is the number of co-rated items that matter.