Design and Analysis of a Gossip-based Trust Recommender System

Stefan Magureanu, Nima Dokoohaki,

Shahab Mokarizadeh and Mihhail Matskin

{magur,nimad,shahab,misha}@kth.se

KTH Royal Institute of Technology, Sweden

KTH VETENSKAP OCH KONST

ROYAL INSTITUTE OF TECHNOLOGY

Introduction

- Prominently used in online stores and social networks
- Emerging interest from Big Data Analysis
- Two main approaches:
  - Neighborhood-based Collaborative Filtering(CF)
  - Matrix Factorization-based(MF)
- Important setback : *Scalability*

ROYAL INSTITUTE
OF TECHNOLOGY

Background - Trust

- Trust propagation:
  - Rely on a pre-defined network of (*social*) trust
  - Compute trust between nodes not directly connected using propagation over existing edges
- Trust inference from ratings:
  - Computed solely from user ratings
  - Usually associated with similarity - not quite

ROYAL INSTITUTE OF TECHNOLOGY

Background - Gossip Algorithms

- T-Man:
  - Network overlay construction algorithm
  - Backbone of our system
  - Uses Cyclon to increase convergence speed
  - Versatile
- Basic Mechanism:
  - Nodes maintain neighborhoods
  - Iteratively exchange neighborhood information with a chosen neighbor
  - Keep most *useful* nodes in neighborhood as described by a *utility function*(or *distance function*)

ROYAL INSTITUTE
OF TECHNOLOGY

Approach - Steps

- Reorganize network, clustering similar users together
- Compute *trust* values towards neighbors
- Improve coverage if needed - recurrent predictions

ROYAL INSTITUTE OF TECHNOLOGY

Approach - T-Man Distance Metric

- Goal:
  - Gather similar users in the neighborhood
  - Use the T-Man neighborhood as the "filter" in CF
- Intuitive metric: Pearson Correlation

$$Similarity(u_1, u_2) = \frac{\sum_i r_{u_1,i} \times r_{u_2,i}}{\sqrt{\sum_i r_{u_1,i}^2} \times \sqrt{\sum_i r_{u_2,i}^2}}$$

- Drawback:
  - Does not take into account number of items in common
  - Users with more rated items in common are more likely to be interested in same items and have rated items the active user is interested in

ROYAL INSTITUTE OF TECHNOLOGY

Approach - T-Man Distance Metric

- Our variation of the Pearson Correlation:

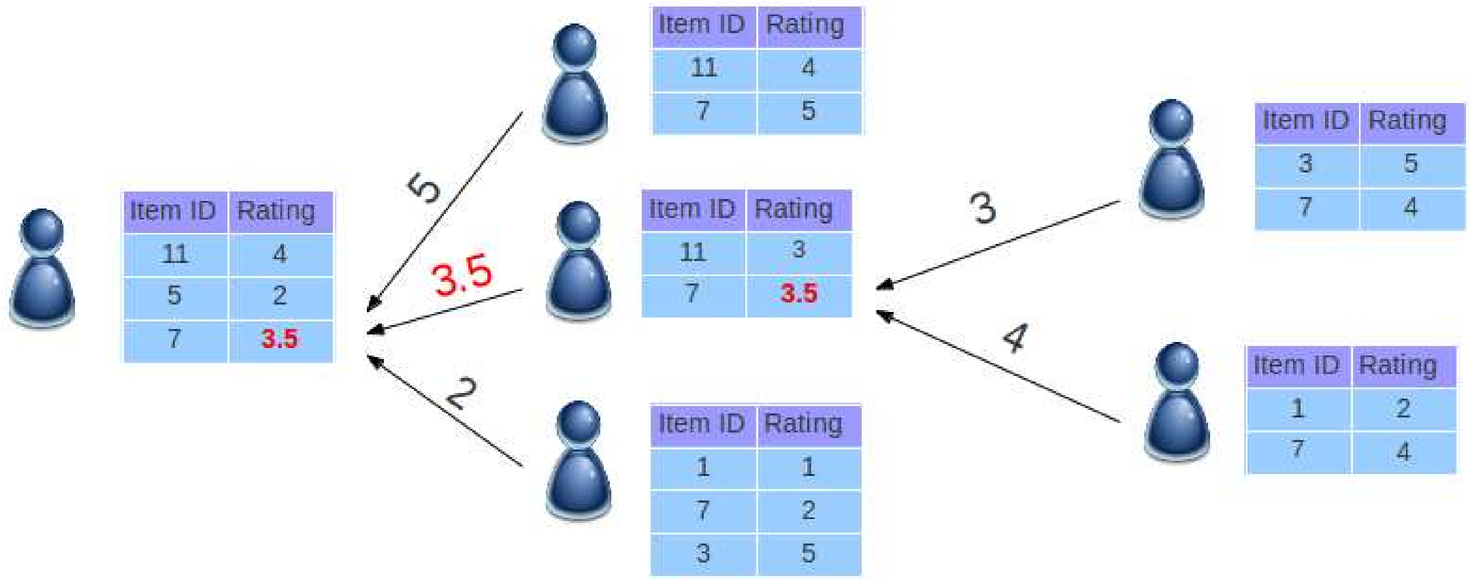
$$Similarity(u_1, u_2) = \frac{\sum_i r_{u_1,i} \times r_{u_2,i}}{\sqrt{\sum_i C \times r^2_{u_1,i}} \times \sqrt{\sum_i r^2_{u_2,i}}}$$

  where $C = 1$ if u_2 rated i else $C = 0.5$.

- Now:
  - Metric penalizes users proportionally to the number of items rated by the active user that they have not rated
  - Compromise between similarity of ratings and similarity of items rated

ROYAL INSTITUTE OF TECHNOLOGY

Approach - Dealing with sparsity



ROYAL INSTITUTE OF TECHNOLOGY

Approach - Modeling trust

Basic idea:

- Use the neighbors to make predictions on the active user's rated item
- Obtain a system of equations with the trust values being the variables
- Solve the system to discover the trusts towards each neighbor

ROYAL INSTITUTE OF TECHNOLOGY

Approach - Modeling trust

- For the prediction formula:

$$\frac{\sum_n (r_{n,i} - \overline{r_n}) \times w_n}{\sum_n w_n} = r_i - \overline{r}$$

- The resulting system would be:

$$\begin{cases} \sum_n \left(r_{n,0} - \overline{r_n} - r_0 + \overline{r}\right) \times w_n = 0 \\ \dots \\ \sum_n \left(r_{n,i} - \overline{r_n} - r_i + \overline{r}\right) \times w_n = 0 \end{cases}$$

- Problem:
  - Solutions must be positive
  - We require positive coefficients to approximate positive solutions

ROYAL INSTITUTE OF TECHNOLOGY

Approach - Modeling trust

- Convert the system to a more useful form by adding a new equation:

$$\begin{cases} \sum_n (r_{n,0} - \overline{r_n} - r_0 + \overline{r}) \times w_n = 0 \\ \ldots \\ \sum_n (r_{n,i} - \overline{r_n} - r_i + \overline{r}) \times w_n = 0 \\ \sum_n w_n = N \times Trust_{mean} \end{cases}$$

  where N is the number of variables and $Trust_{mean}$ is the average desired trust value.

ROYAL INSTITUTE OF TECHNOLOGY

Approach - Modeling trust

- Obtain positive coefficients:

$$\begin{cases} \sum_n (2 \times R_{max} + r_{n,0} - \overline{r_n} - r_0 + \overline{r}) \times w_n = C \\ \ldots \\ \sum_n (2 \times R_{max} + r_{n,i} - \overline{r_n} - r_i + \overline{r}) \times w_n = C \\ \sum_n w_n = N \times Trust_{mean} \end{cases}$$

  where $C = 2 \times R_{max} \times \sum_n w_n$ and N is the number of variables.

- We can now chose a convex method to approximate the solutions

ROYAL INSTITUTE OF TECHNOLOGY

Approach - Approximating trust values

- Our solving algorithm is proposed by D. Cartwright in 2011
- Uses Expectation Maximization to converge to approximations of positive solutions
- Can be reduced to the following form:

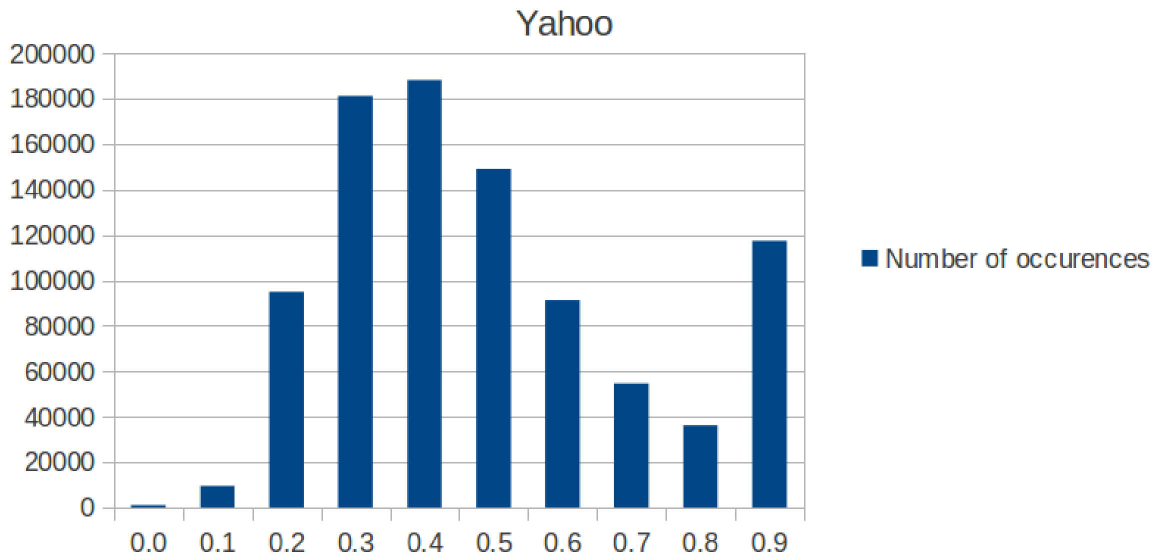
```
while(!done)
    for each n in neighbors
```

$$w_n = w_n \times \frac{\sum_{i \in Items} \frac{C}{C_i} \times coef_{i,n}}{\sum_{i \in Items} coef_{i,n}}$$

ROYAL INSTITUTE OF TECHNOLOGY

Approach - Approximating trust values

Obtained Trust Distribution:



ROYAL INSTITUTE OF TECHNOLOGY

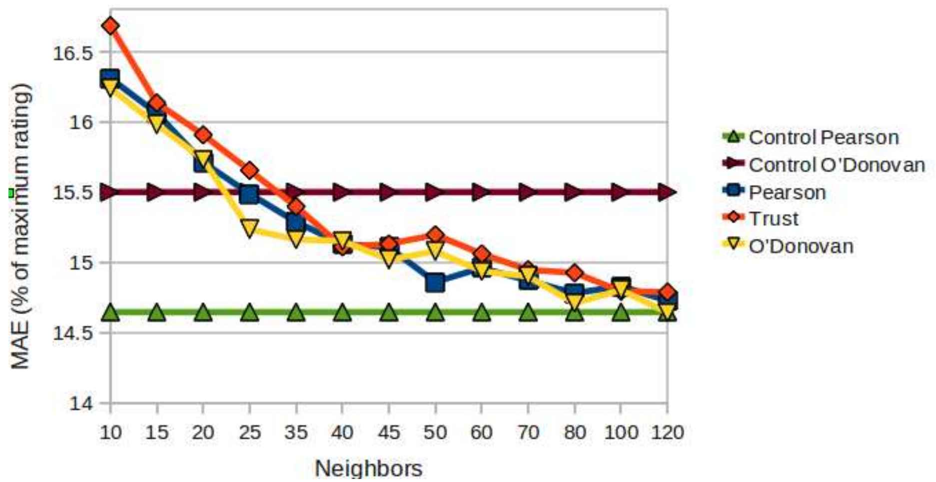
Results - Setup

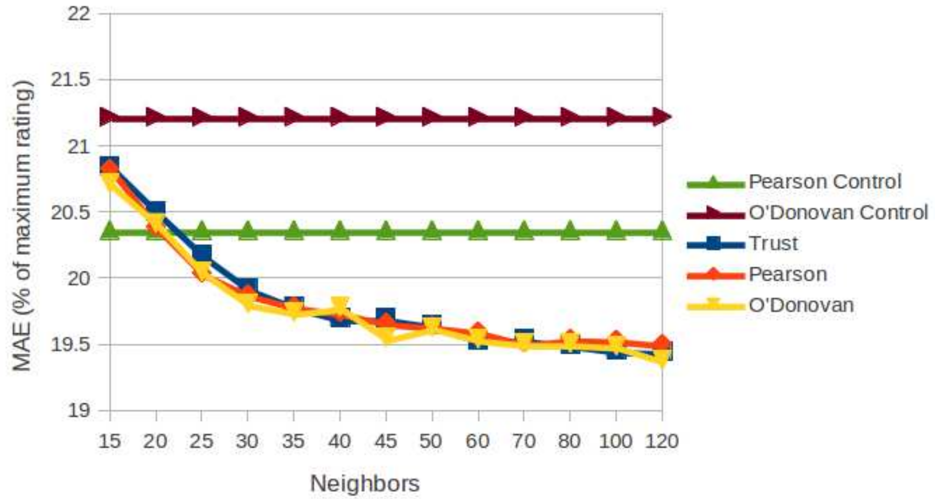
- Leave one out
  - Hide random item from each user
  - Run T-Man for X rounds
  - Predict hidden item ratings
  - Coverage = % of hidden items for which the system could generate a prediction
- 3 Datasets
  - Movielens - 6,040 users, 1,000,000 ratings, 4,000 items
  - Yahoo Webscope! - 15,400 users, 300,000 ratings, 1,000 items
  - Epinions - 49,290 users, 664,824 ratings, 140,000 items

ROYAL INSTITUTE
OF TECHNOLOGY

Results - MAE

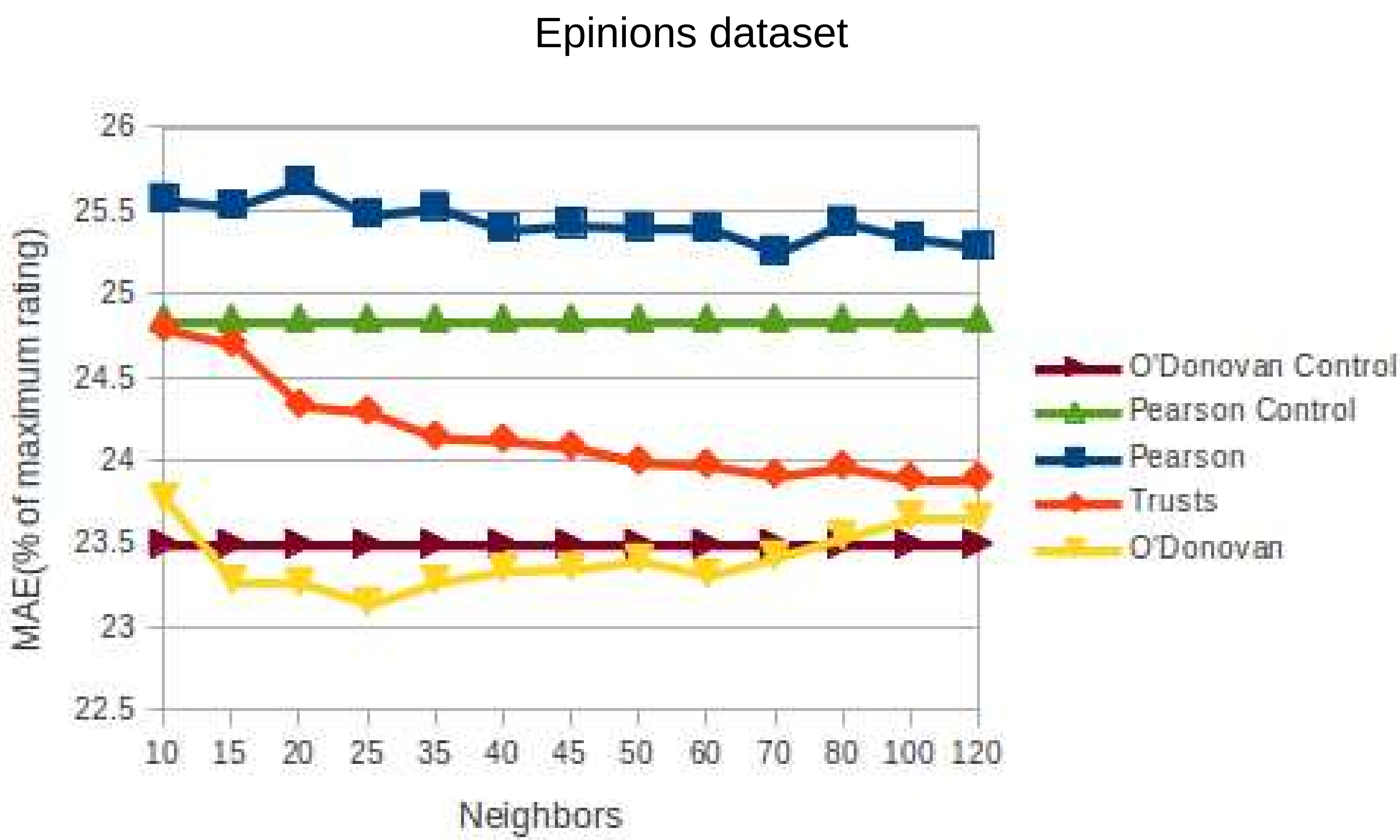
Movielens(top) and Yahoo(bottom) datasets.



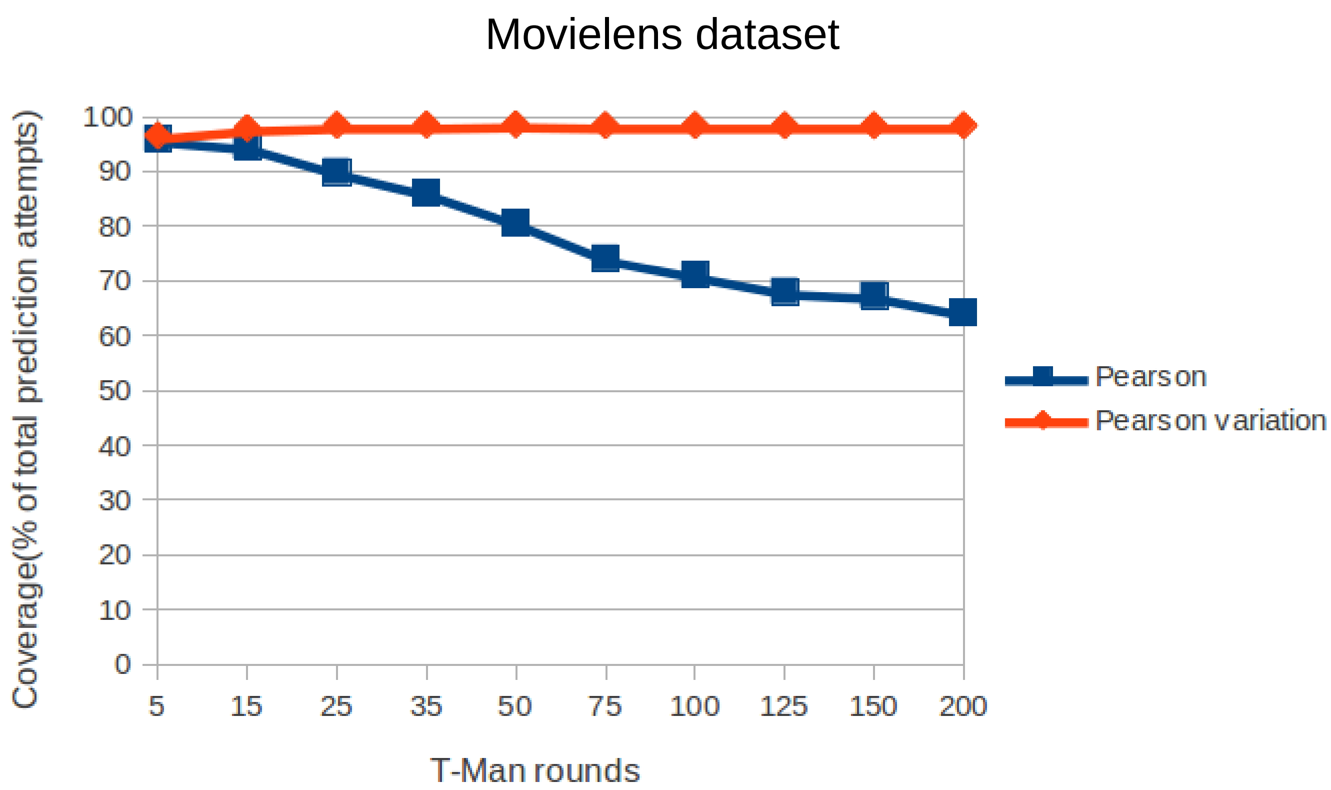


ROYAL INSTITUTE OF TECHNOLOGY

Results - MAE

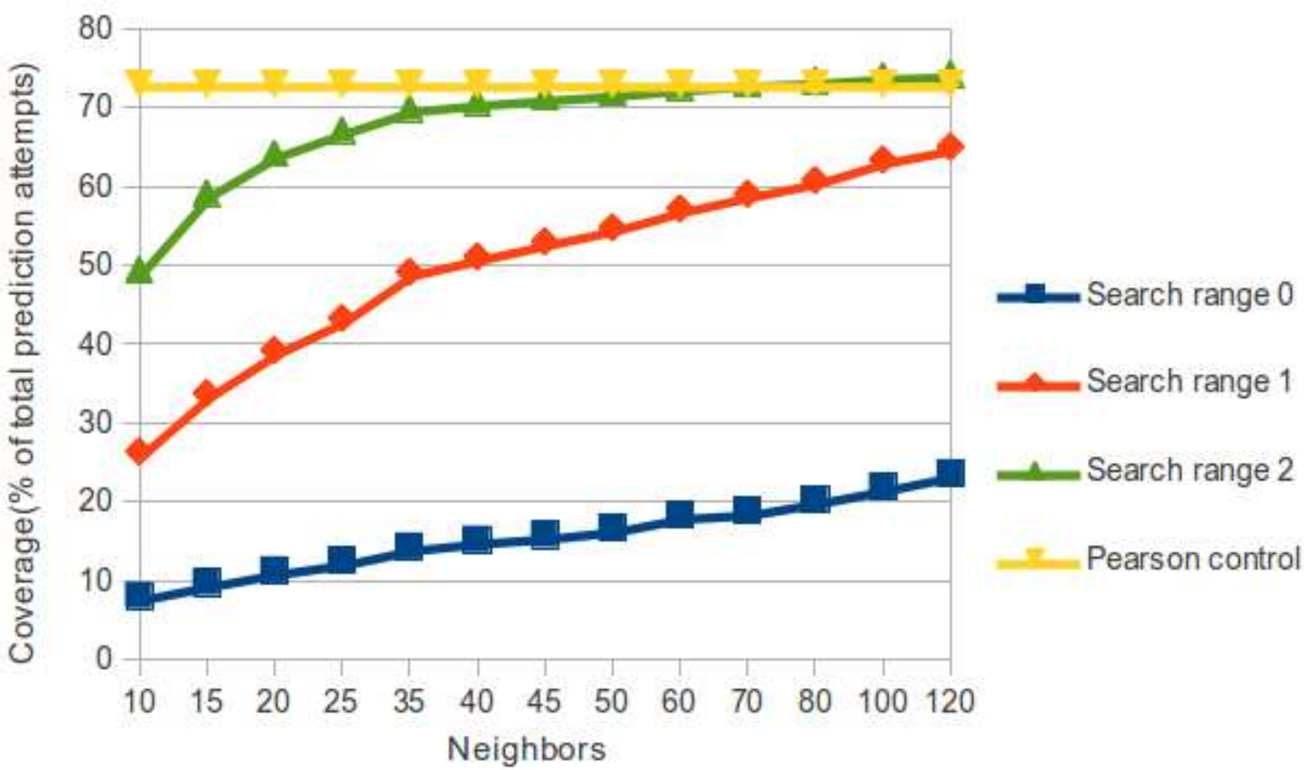

Epinions dataset
26
25.5
25
24.5
24
23.5
23
22.5
MAE(% of maximum rating)
10
15
20
25
35
40
45
50
60
70
80
100
120
Neighbors
O'Donovan Control
Pearson Control
Pearson
Trusts
O'Donovan

Results - Coverage



Results - Coverage Search Range

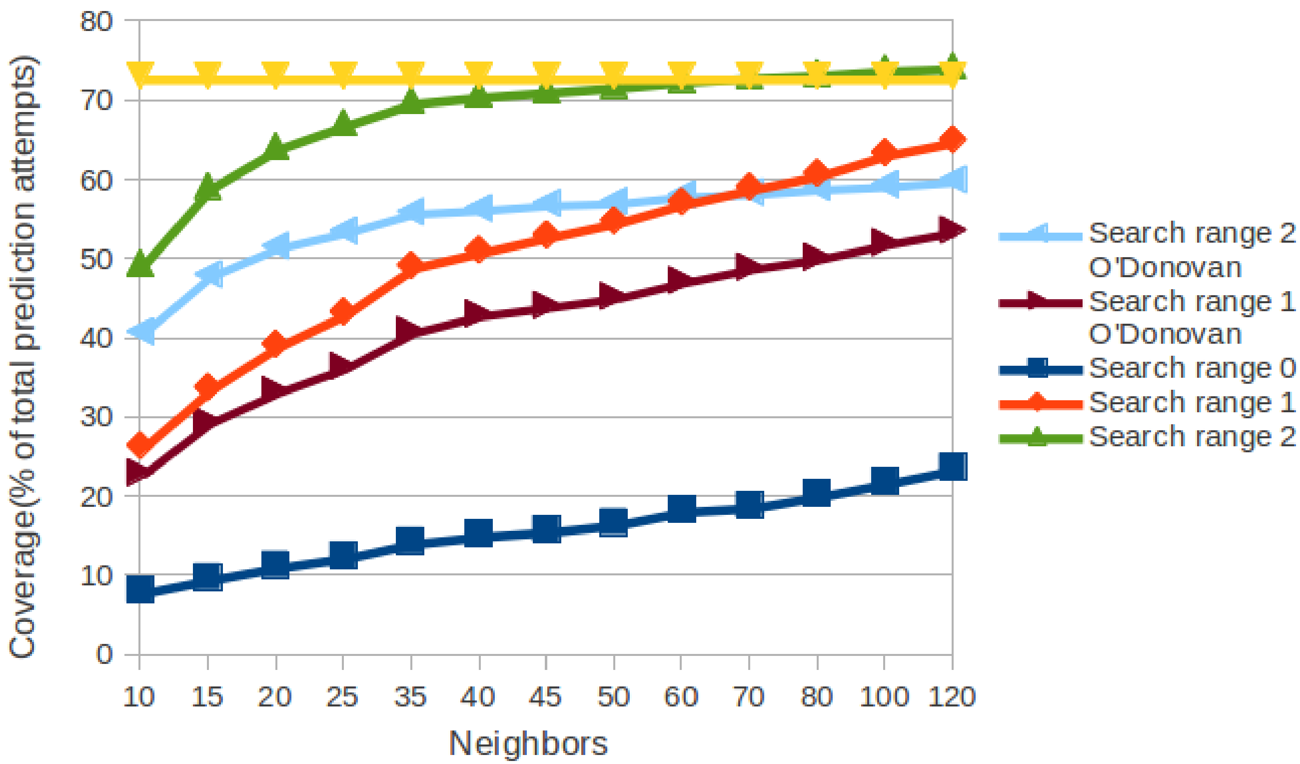
Epinions dataset - effect of search range on coverage.



ROYAL INSTITUTE OF TECHNOLOGY

Results - Coverage
Search Range

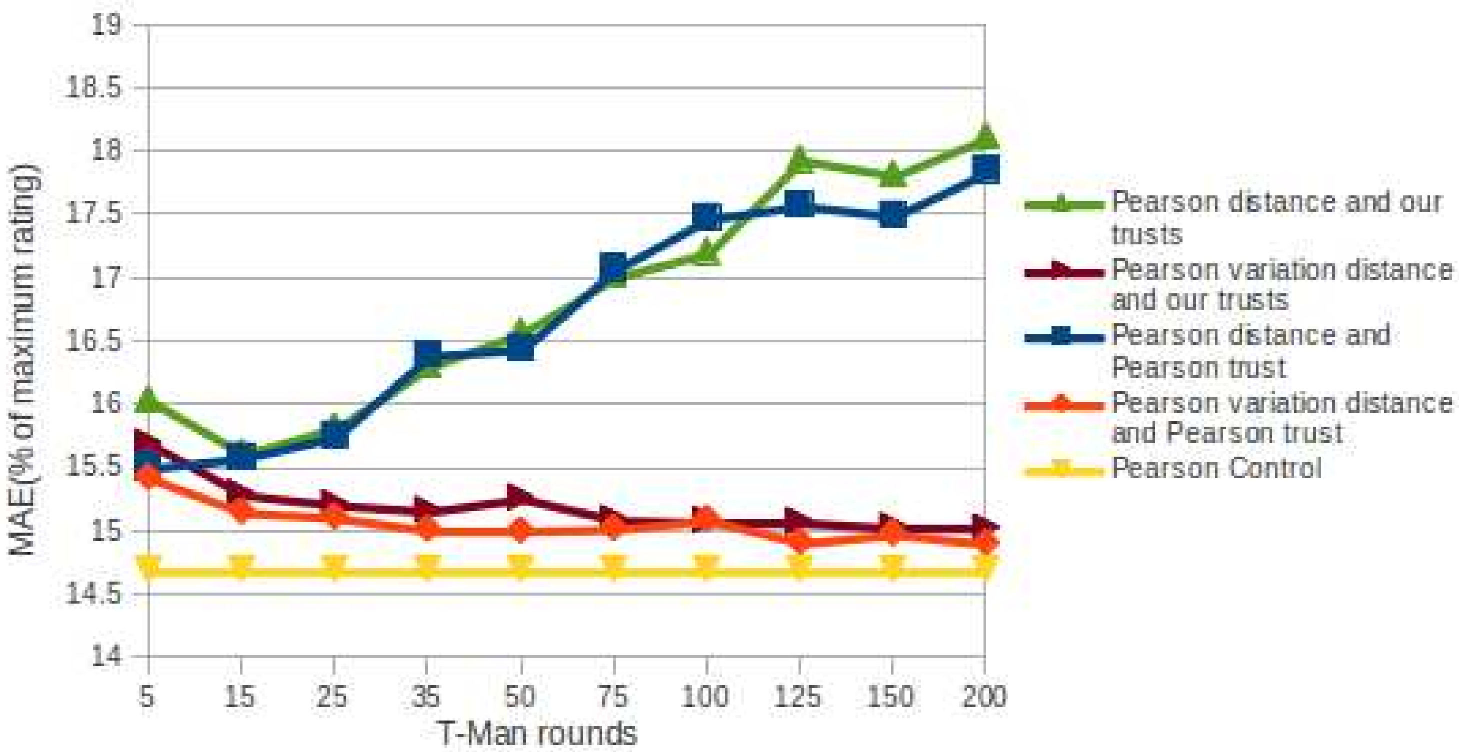
Epinions dataset - effect of search range on coverage.



ROYAL INSTITUTE OF TECHNOLOGY

Results - Coverage Effects on MAE

Movielens dataset - effect of distance metric on MAE.



ROYAL INSTITUTE OF TECHNOLOGY

Conclusions

- Proposed a fully decentralized CF recommender system
- Presented methods of increasing coverage
- Proposed new trust inference model
- Future work:
  - Evaluate diversity of recommendation
  - Evaluate trust models based on boosting
  - Implement item-relevance
  - Model trust distribution in the system

ROYAL INSTITUTE OF TECHNOLOGY

Questions?

Contact information :

Stefan Magureanu, magur@kth.se

ROYAL INSTITUTE OF TECHNOLOGY