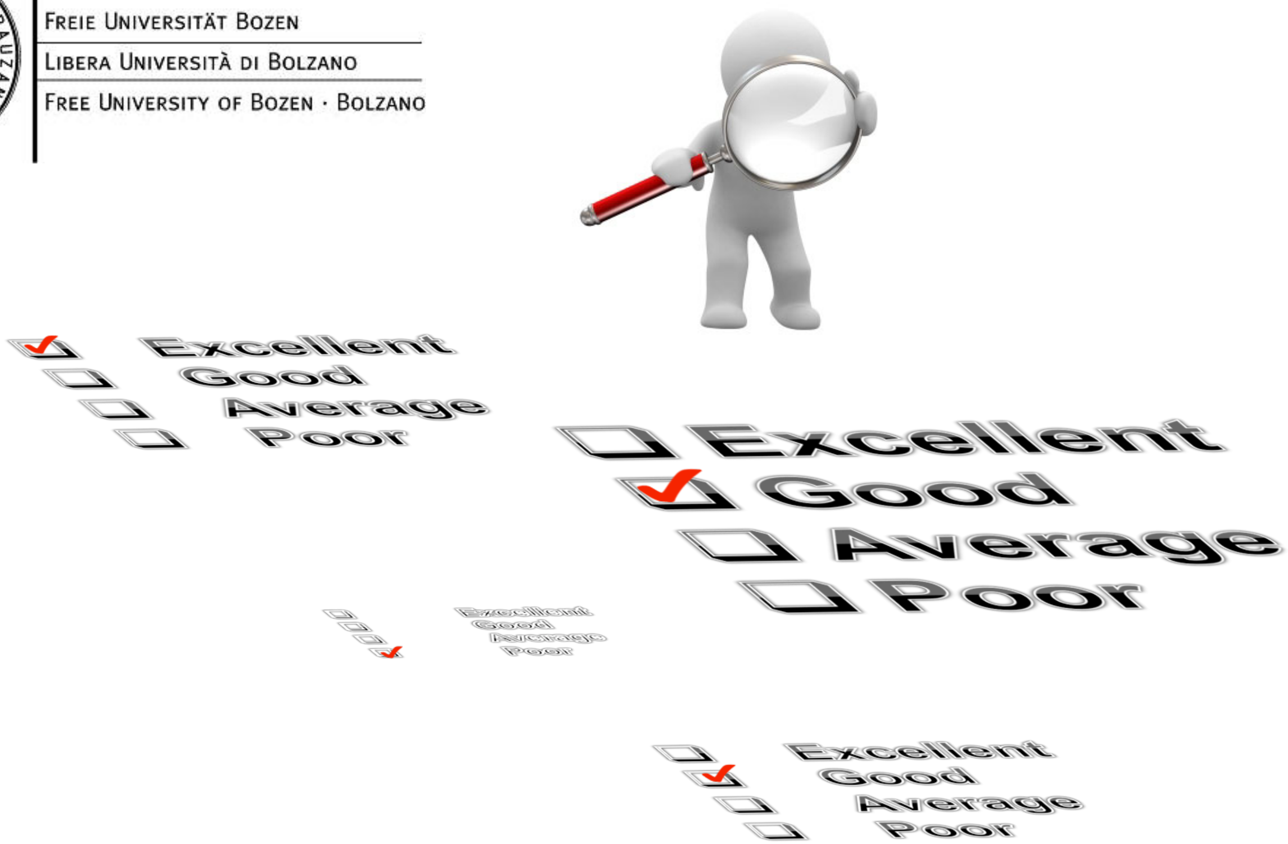
Mehdi Elahi
Francesco Ricci
Valdemaras Repsys

Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano



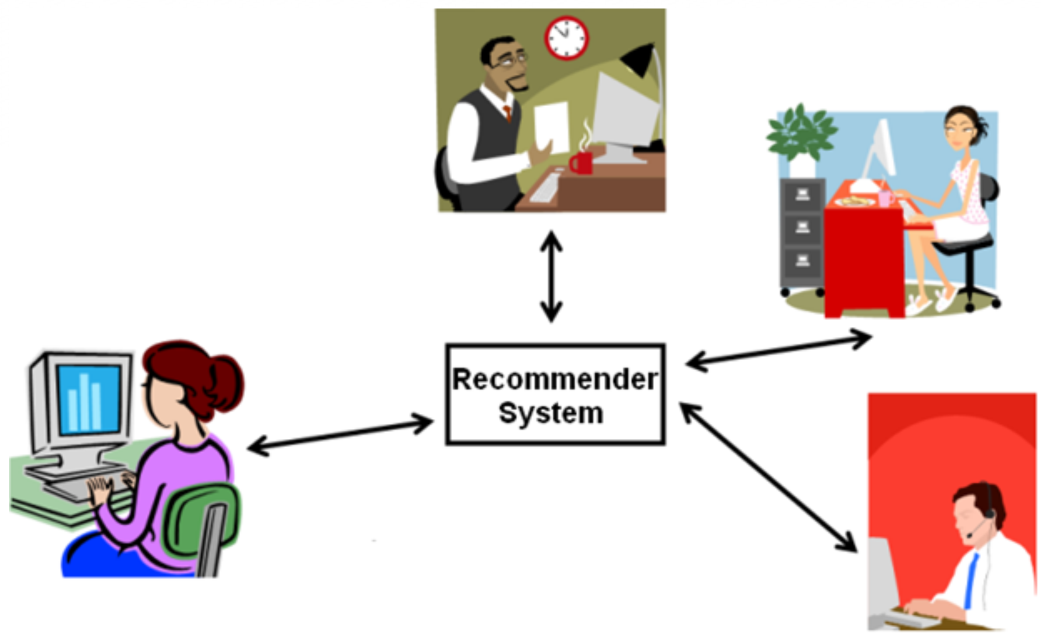
IJCAI

System-Wide Effectiveness of Active Learning in Collaborative Filtering

Free University of Bolzano
Bolzano, Italy

International Workshop on Social Web Mining co-located with **IJCAI 2011**, Barcelona, Spain

INTRODUCTION



- **Collaborative Filtering:**
- A technique used to predict the ratings exploiting ratings given by the users to items.

Sparsity of the Data

- In Netflix: **98.8 %**

  of the data is unknown

- In Movielens: **95.7 %**

  of the data is unknown

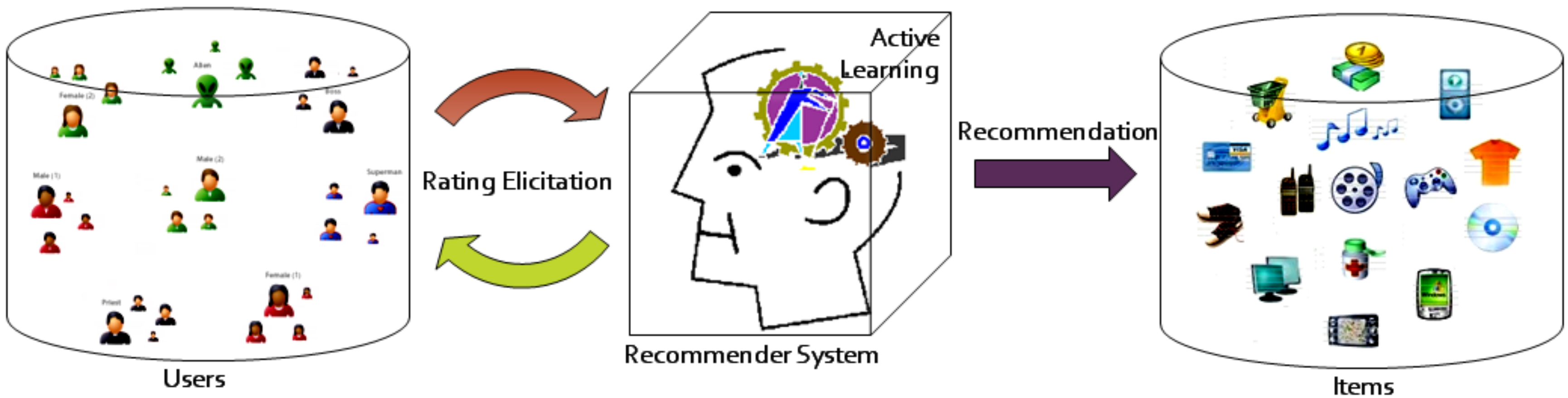
Items

Users

	1	2	3	4	5	6	7	8	9
1							2		
2		5						3	1
3				1	5				
4	5							5	
5									
6	3			1			4		
7									
8		5	4						
9							5		

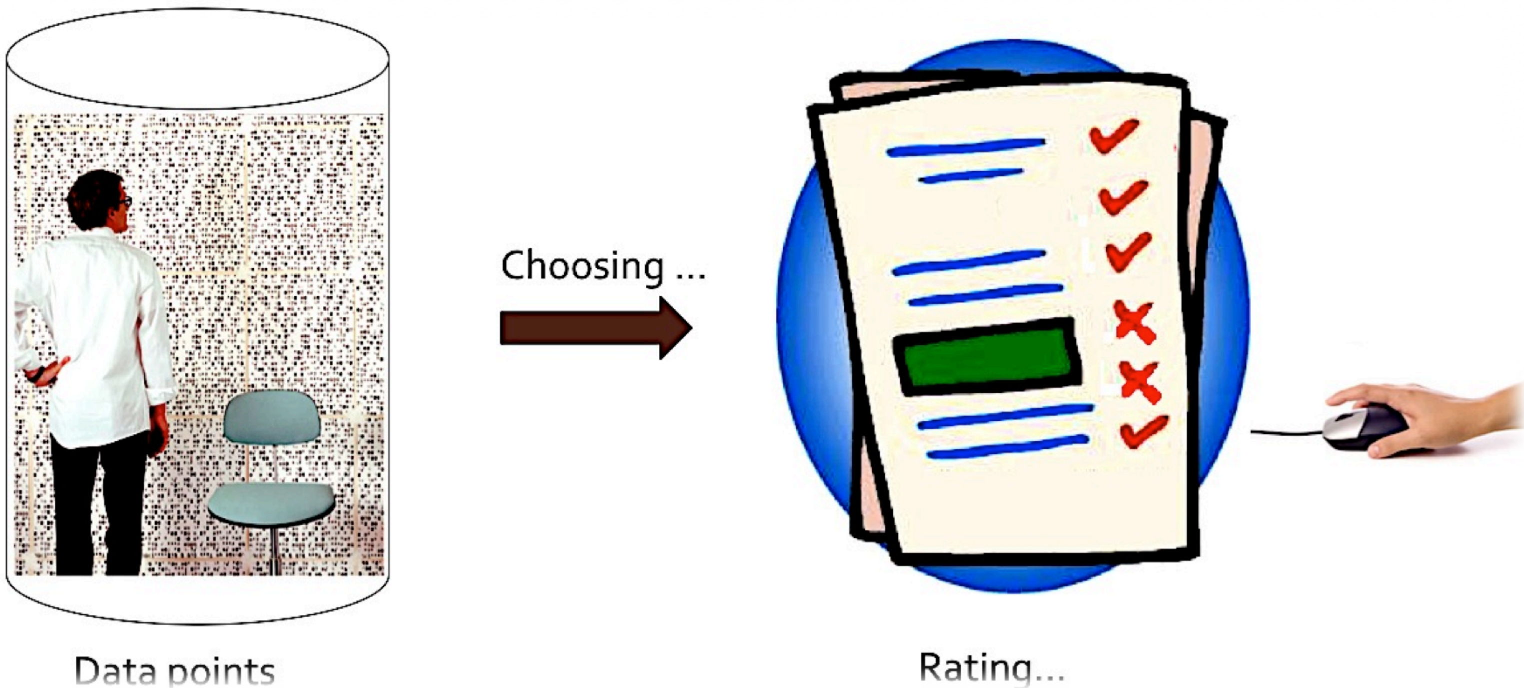
Ratings

Active Learning with Collaborative Filtering



- **Active Learning:**
- If during the learning process – rating prediction - some preferences are not available, the system can actively and selectively ask the user their value

Definition of the Active Learning Strategy



- An active learning **strategy** for a Collaborative Filtering is a set of rules to choose the best items for the users to rate.

Objectives

- **To simulate the evolution of a RS and its performance** exploiting rating elicitation strategies, i.e., algorithms for choosing the items to be presented to the user for rating
- **To understand the benefits and drawbacks of different strategies** with respect to various measures of a recommender system effectiveness (e.g. Mean Absolute Error, precision, ranking quality, or coverage)
- **To study whether the rating elicitation strategy must take into account the size** and the state of the rating database.

Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

Some State-of-the-art Strategies

- **Popularity:** chooses the most popular items, hence it is more likely that a request of such a rating will really increase the size of the rating database (Carenini, 2003)

- **Entropy:** items with highest entropy value are chosen which have diverse ratings (Rashid, 2002 and 2008)

- **Variance:** collects the opinion of the user on items with more diverse ratings - assuming that these ratings are more useful (Rubens, 2011)

- **log(Popularity)*Entropy:** combines the popularity with entropy strategy; selected popular items with higher entropy (Rashid, 2002 and 2008)

- All the related works have focused on the **single user problem,** but We focus on **system-wide effectiveness**

Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

Additional Strategies

- **Highest Predicted:** the items with largest predicted ratings are more likely to have been experienced by the user and their true ratings reveal what the user likes
- **Lowest Predicted:** reveals what the user does not like, but can actually collect a few ratings, since the user is unlikely to have experienced all the items that he does not like
- **Highest-Lowest Predicted:** combines "highest predicted" and "lowest predicted" strategies
- **Binary Prediction:** tries to predict what items the user has experienced, to maximize the probability that the user be able to rate the item
- **Voting**: A committee of strategies vote for items and the most voted items are chosen
- Prediction based strategies try to select items that the **user is likely to be able to rate**

Evaluation Methodology

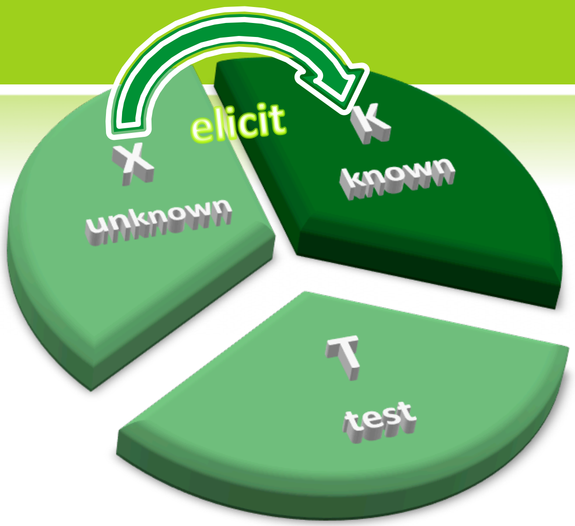
Movielens

No. of users: 943
No. of items: 1682
No. of ratings: 100K

Netflix

No. of users: 480189
No. of items: 17770
No. of ratings: 100M*

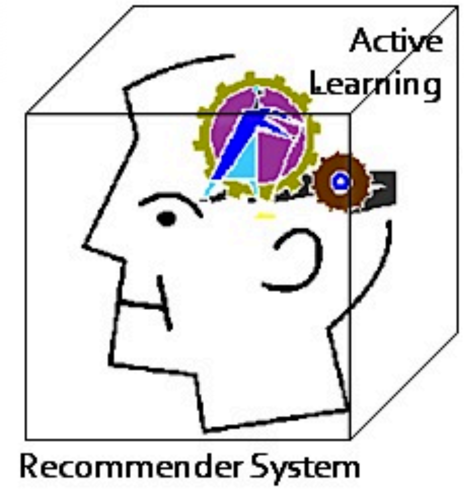
* We used 1st 100K



- **Datasets are partitioned into three subsets:**
- **Known (K):** contains the rating values that are considered to be known by the system at a certain point in time.
- **Unknown (X):** contains the rating values that are considered to be known by the users but not to the system. These ratings are incrementally elicited, i.e., their values are transferred into **K** if the system asks them to the (simulated) users.
- **Test (T):** contains the ratings that are never elicited and are used only to test the recommendation effectiveness after the system has acquired the new elicited ratings.

Learning Iteration

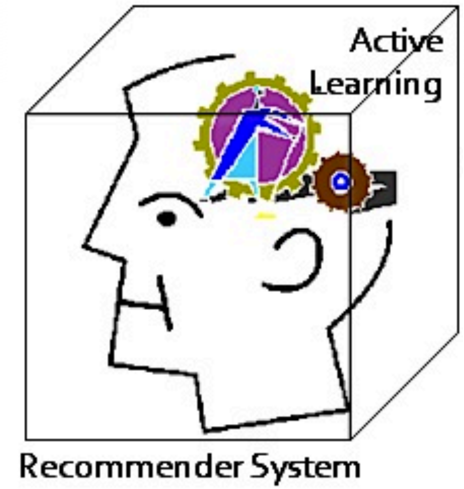
Item	Score
1	5
2	4
3	1
4	1
5	4
6	4
7	3
8	5
9	2
…	…
N	1



System computes the score for all the items that can be scored (according to a strategy)

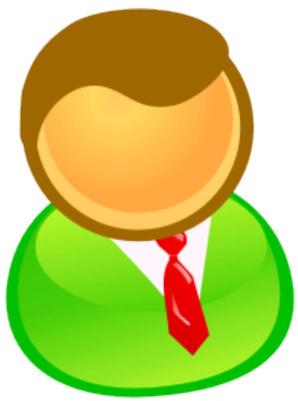
Learning Iteration

Item
1
2
3
4
5
6
7
8
9
…
N

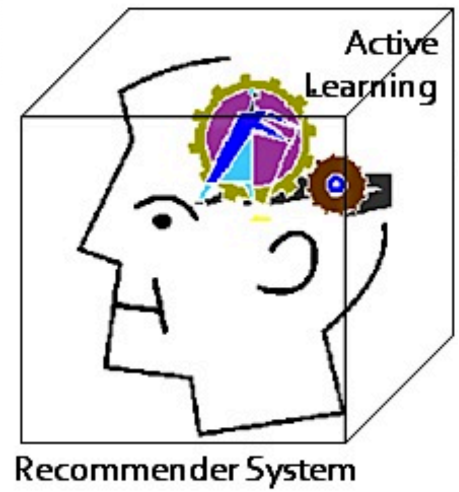


The system selects the top 10 items and presents them to the simulated user

Learning Iteration

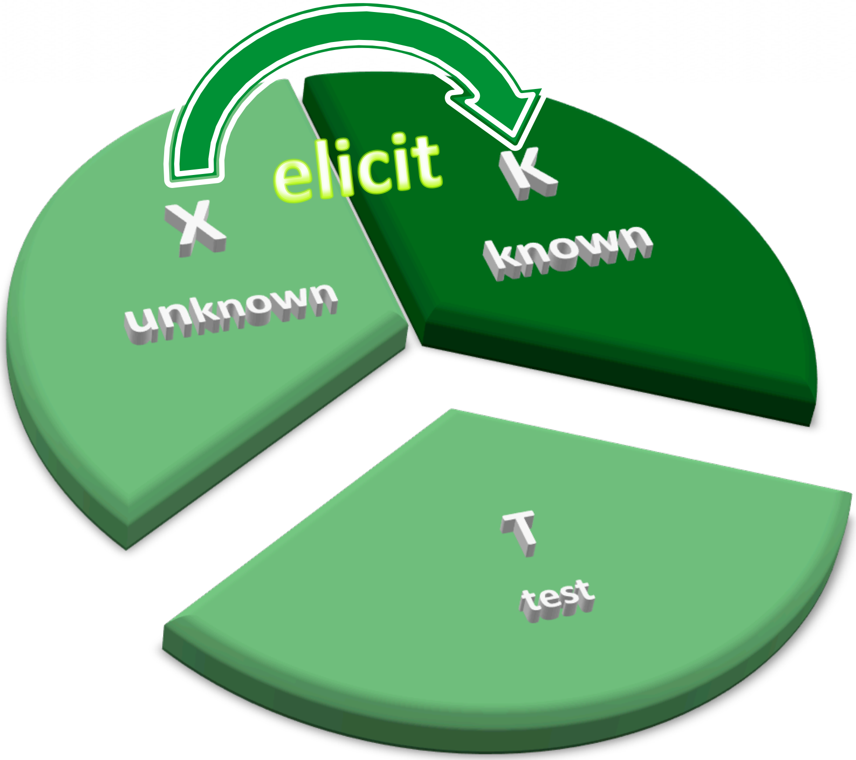


Rated items
1
2
5
75
13



The items that are rated in the unknown set (X) are found and transferred to the known set (K)

Learning Iteration

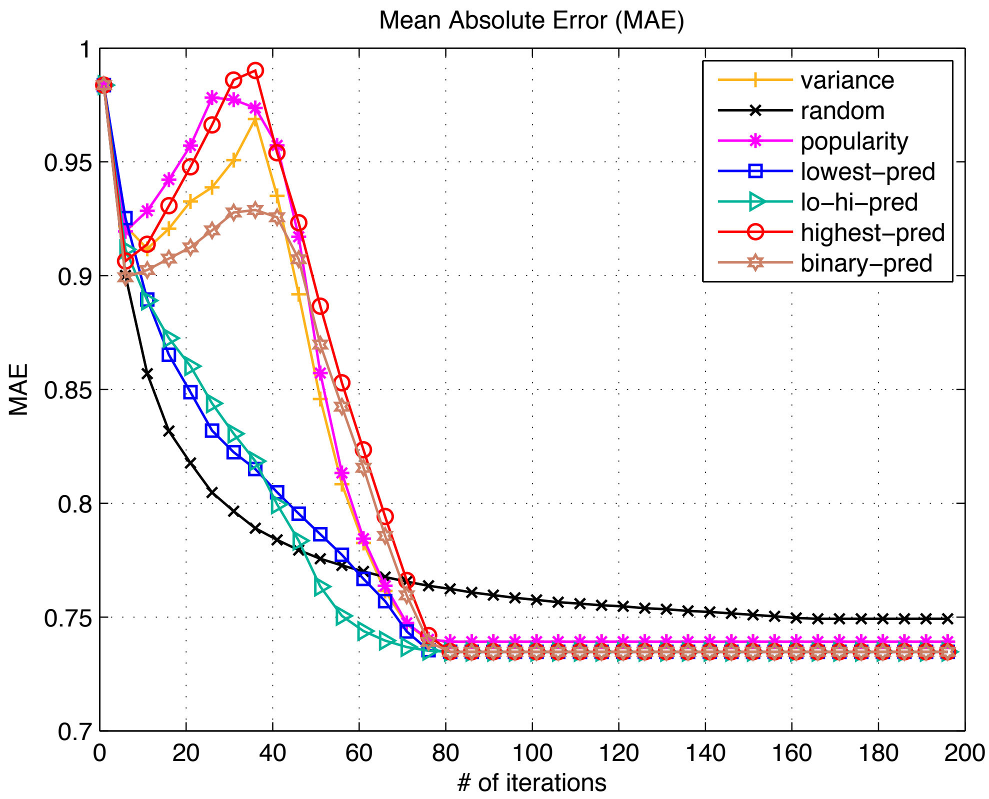


The items that are rated in the unknown set (X) are found and transferred to the known set (K)

Evaluation: MAE

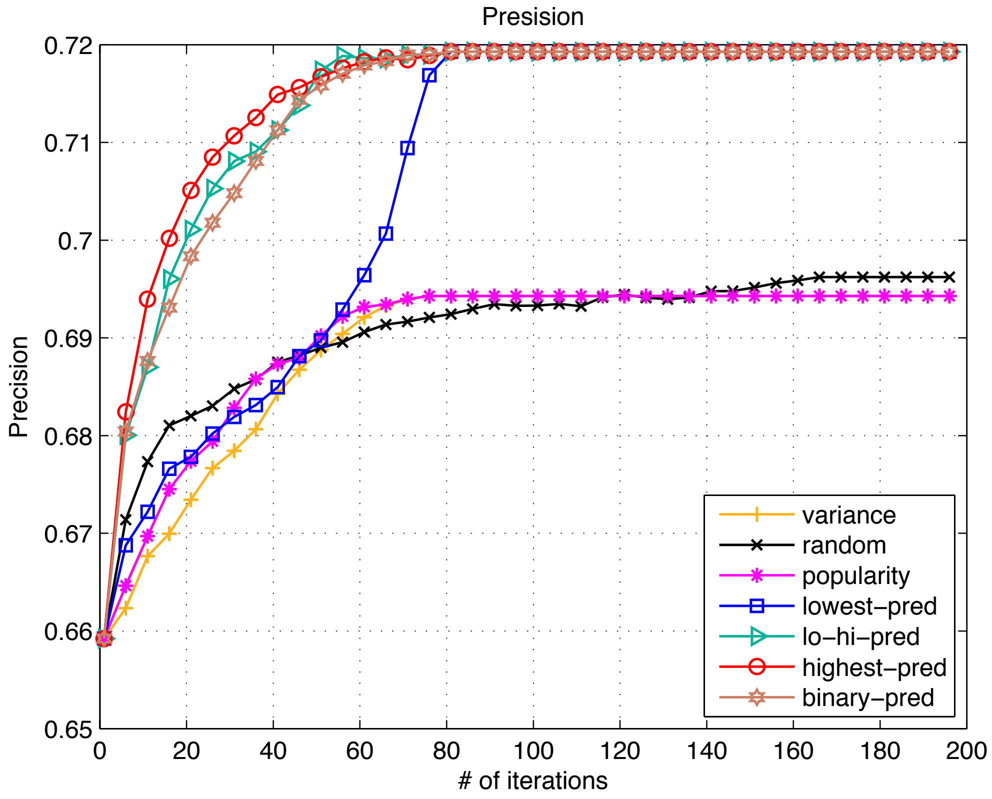
- **Mean Absolute Error (MAE)**
- Measures the average absolute deviation of the predicted rating from the user's true rating:

$$MAE = \frac{1}{|T|} \sum_{r_{u,i} \in T} |r_{u,i} - \hat{r}_{u,i}|$$



Evaluation: Precision

- **Precision:** percentage of the items with rating values (as in ***T***) equal to 4 or 5 in the top 10 recommended items.

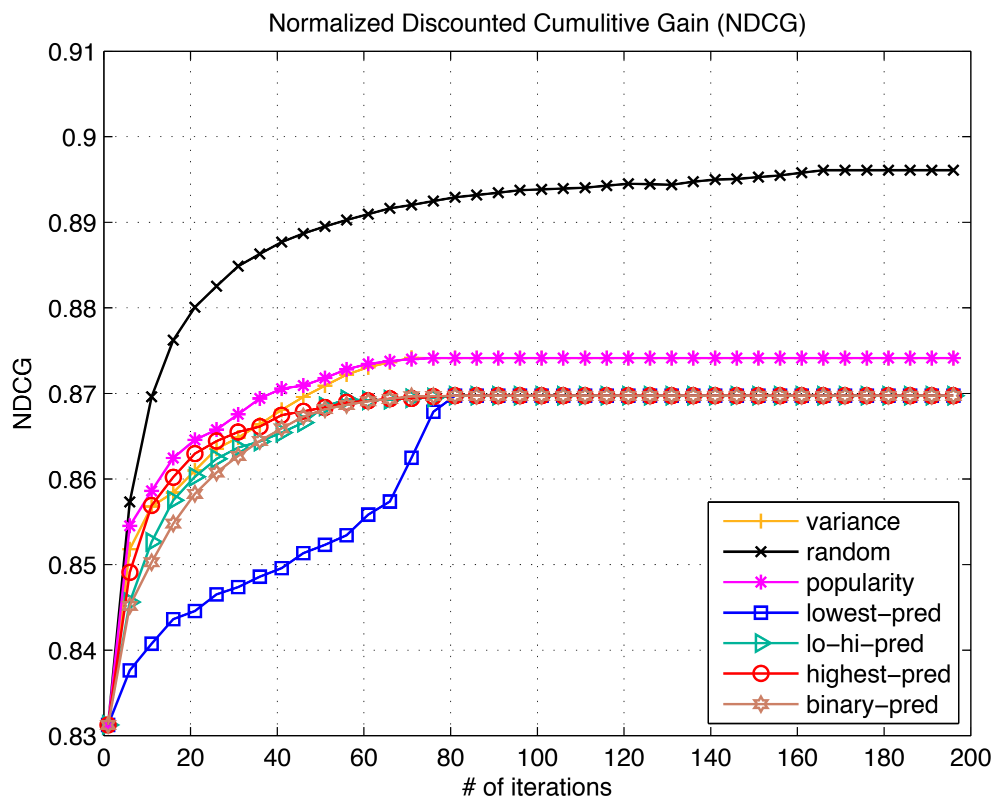


Evaluation: NDCG

- **Normalized Discounted Cumulative Gain (NDCG):**
- The recommendations for ***u*** are sorted according to the predicted rating values, then **DCGu** is defined as:

$$DCG_u = r_u^1 + \sum_{i=2}^{N} \frac{r_u^i}{log_2 i}$$

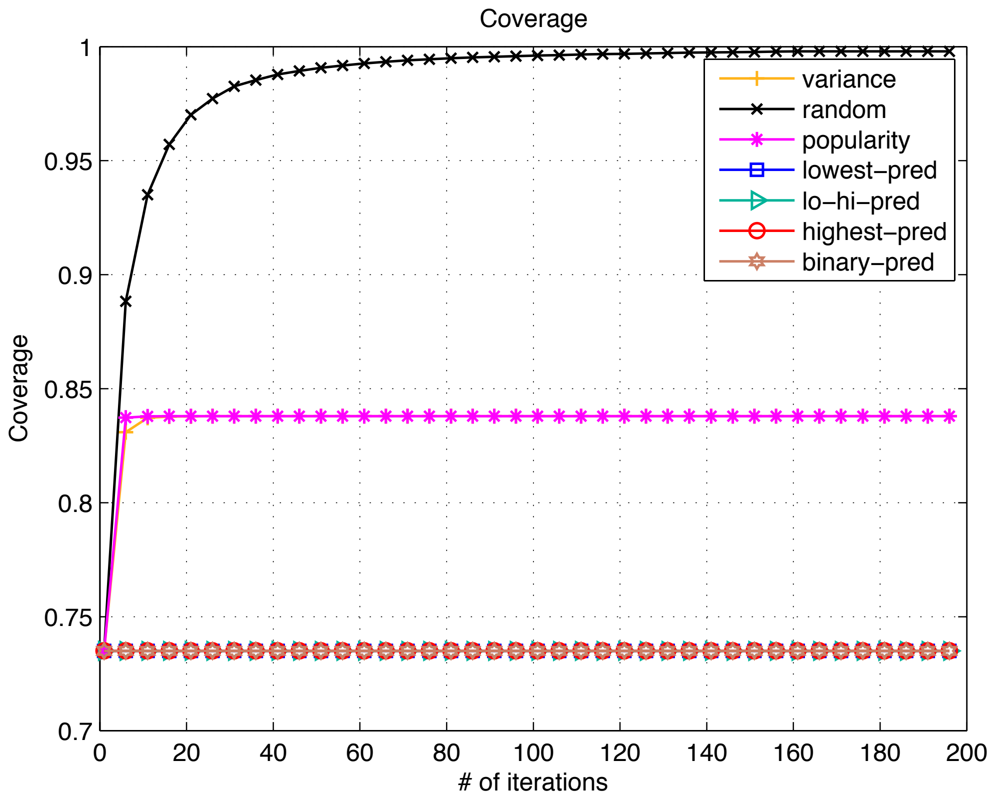
$$NDCG_u = \frac{DCG_u}{IDCG_u}$$



Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

Evaluation: Coverage

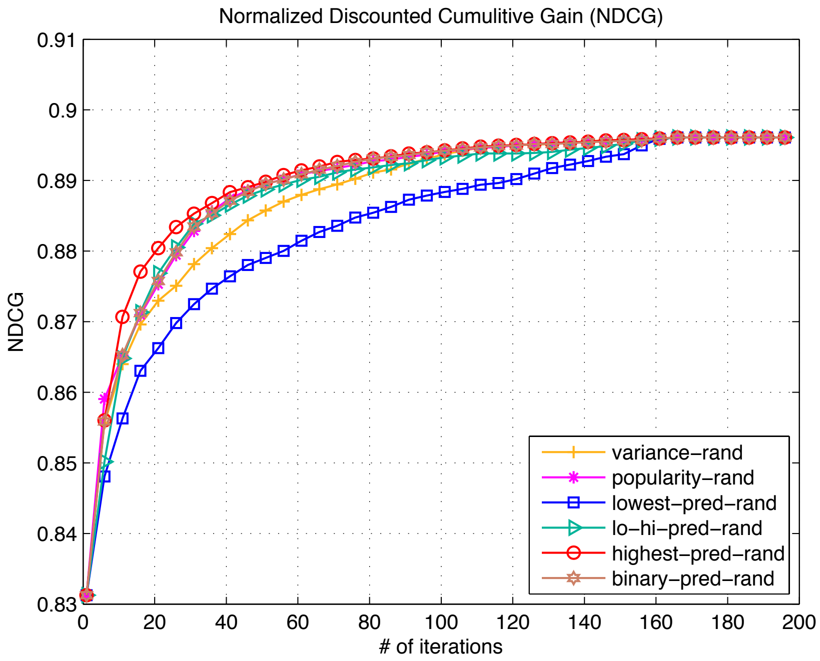
- **Coverage:** proportion of the full set of items over which the system can form predictions or make recommendations.

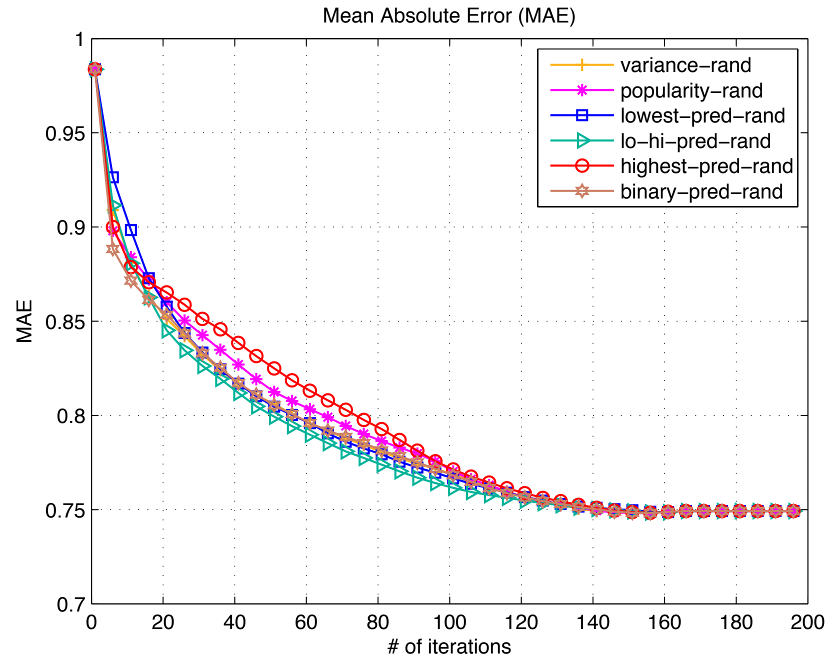


Partially-randomized Strategies

- The list of items returned by a pure strategy S is modified, **introducing some random items**, simulating the free addition of some rating values not explicitly requested by the system but known by the user.
- For instance, we note that if ***S*** is the highest predicted strategy, there are **cases where no rating predictions can be computed** by the RS for the user ***u***, and hence ***S*** would not be able to identify the items to be rated.
- This happens when ***u*** is a new user and none of his ratings are known. In this case the **randomized version** of this strategy generates purely random items for the user to rate.

Evaluation: MAE and NDCG





Conclusion

- We have made new experimental design considering **system-wide effectiveness** of rating elicitation strategies rather than how good they are for a single user
- Our results will help selecting the right strategy for a given effectiveness metric - in fact, **there is no single best strategy**, among those that are evaluated, that dominates the others for all the evaluation measures.
- We have considered several metrics such as **MAE**, **NDCG**, **Precision** ,and **Coverage**
- The **lo-high predicted** is the best for MAE and precision
- The **random** strategy is the best for NDCG

Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

THANK YOU!

FREIE UNIVERSITÄT BOZEN
LIBERA UNIVERSITÀ DI BOLZANO
FREE UNIVERSITY OF BOZEN · BOLZANO