Performance prediction and evaluation in Recommender Systems: An Information Retrieval perspective

Alejandro Bellogín Kouki

under the supervision of

Pablo Castells Azpilicueta

and

Iván Cantador Gutiérrez

Introduction (1)

bing Beta | olympic games

Olympic Games | Vancouver 2010 | London 2012 | Sochi 2014 | Rio 2016
www.**olympic**.org/**olympic-games**
Beijing's iconic new sporting venues provide the backdrop to record-breaking feats!

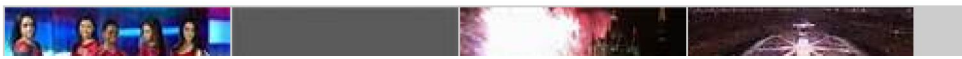
London 2012 **Olympics** - Schedule, Results, Medals, Tickets, Venues
www.london2012.com
[{"id":"1304296","rel":{}},{"id":"1408355","rel":{}},{"id":"1408600","rel ... Official site of the London 2012 **Olympic** and Paralympic **Games**

Images of **olympic games**
bing.com/images

Olympic Games Medals, Results, Sports, Athletes | London 2012 ...
www.**olympic**.org
Official source of **Olympic Games** sports, countries, results, medals, schedule, athlete bios, teams, news, photos, videos for Summer and Winter **Olympics**

Olympic Games - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Olympic_Games**
Ancient **Olympics** · Modern **Games** · International **Olympic** ... · Commercialization
The modern **Olympic Games** (French: les Jeux olympiques, JO) are a major international event featuring summer and winter sports in which thousands of athletes ...

Videos of **olympic games**
bing.com/videos



Google | olympic games

Olympic Games - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Olympic_Games**
The modern **Olympic Games** (French: les Jeux olympiques, JO) are a major international event featuring summer and winter sports in which thousands of **...**
List of Olympic Games host cities - Summer Olympic Games - Olympic sports - 1906

London 2012 **Olympics** - Schedule, Results, Medals, Tickets, Venues
www.london2012.com/
Inspire a generation. **Olympic Games** 27 July - 12 August. Paralympic Games 29 Aug - 9 Sept. Official London 2012 website **...**

Olympic Games | Vancouver 2010 | London 2012 | Sochi 2014 | Rio **...**
www.**olympic**.org/**olympic-games**
Watch videos, photos and news from past and future **Olympic Games**: Beijing 2008, Vancouver 2010, London 2012, Sochi 2014 and Rio 2016.

Ancient **Olympic Games** | History, Mythology, The Athlete, Sport Events
www.olympic.org › Olympic Games
According to historical records, the first ancient **Olympic Games** can be traced back to 776 BC. They were dedicated to Olympian gods and staged in Olympia.

Olympic Games Medals, Results, Sports, Athletes | London 2012 **...**
www.**olympic**.org/
Official source of **Olympic Games** sports, countries, results, medals, schedule, athlete bios, teams, news, photos, videos for Summer and Winter Olympics.

London 2012 Summer **Olympics** | **Olympic** Videos, Photos, News
www.olympic.org › Olympic Games
20+ items – london 2012 **Olympic Games** video highlights, photos, results, **...**
Jade Jones Competing In The Final 8/9/2012
Musicians Welcome Team GB To The **Olympic** Village 7/26/2012

Introduction (2)

Is it possible to anticipate the success of a search before its execution?

Introduction (3)

- In Information Retrieval (IR), **performance prediction techniques** address how to estimate the performance of a **query**
  - In a given collection
  - Based on the collection's vocabulary and statistics
  - Using (or not) the retrieved documents

- We study the performance prediction problem in recommendation
  - Where no query is given

Recommender Systems (1)

- A recommender system aims to find and suggest items of **likely interest** based on the **users' preferences**

- Examples:
  - **Amazon** – products
  - **Netflix** – tv shows and movies
  - **LinkedIn** – jobs and colleagues
  - **Last.fm** – music artists and tracks

Recommender Systems (2)

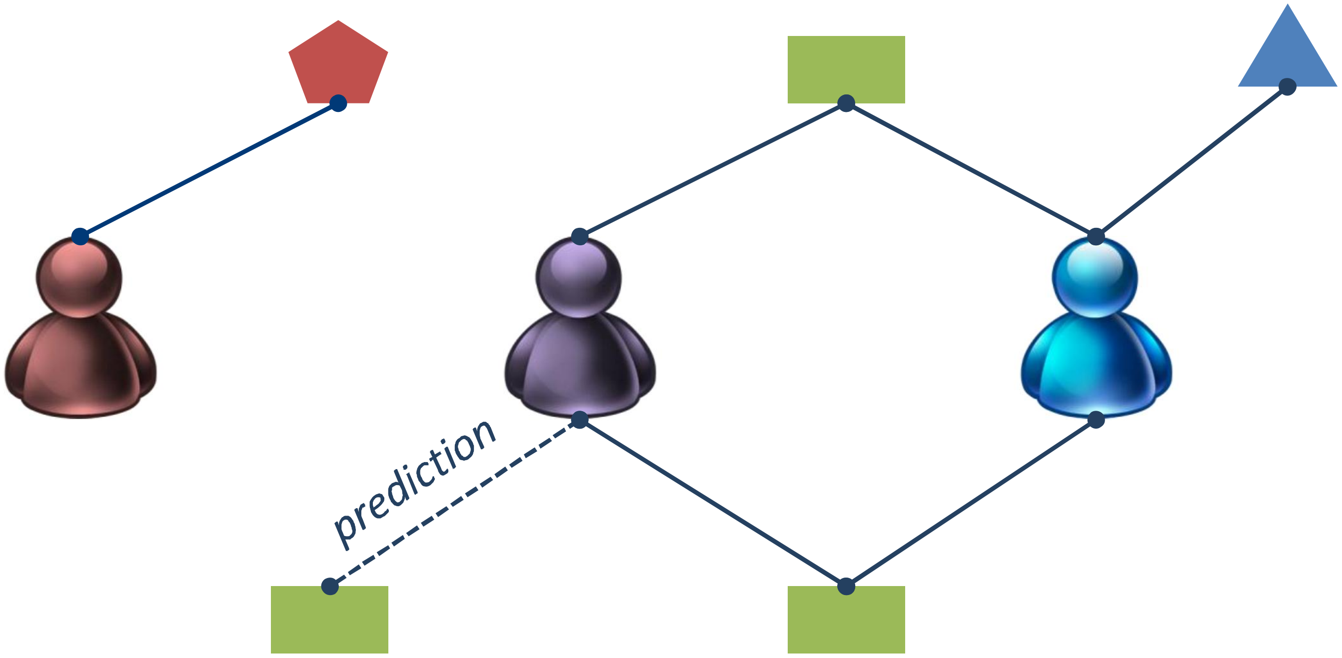
- The interactions between the user and the system are recorded
  - Typically, in the form of ratings

	i_1	…	i_k	…	i_m
u_1	★★★★★		★★★★★		★☆☆☆☆
⋮					
u_j	★★★☆☆		?		★★☆☆☆
⋮					
u_n	★★☆☆☆		★★★★☆		★★☆☆☆

- The items could be of any type: movies, music, people, …

Recommender Systems (3)

- Item suggestions can be obtained using several techniques:
  - **Content-based**

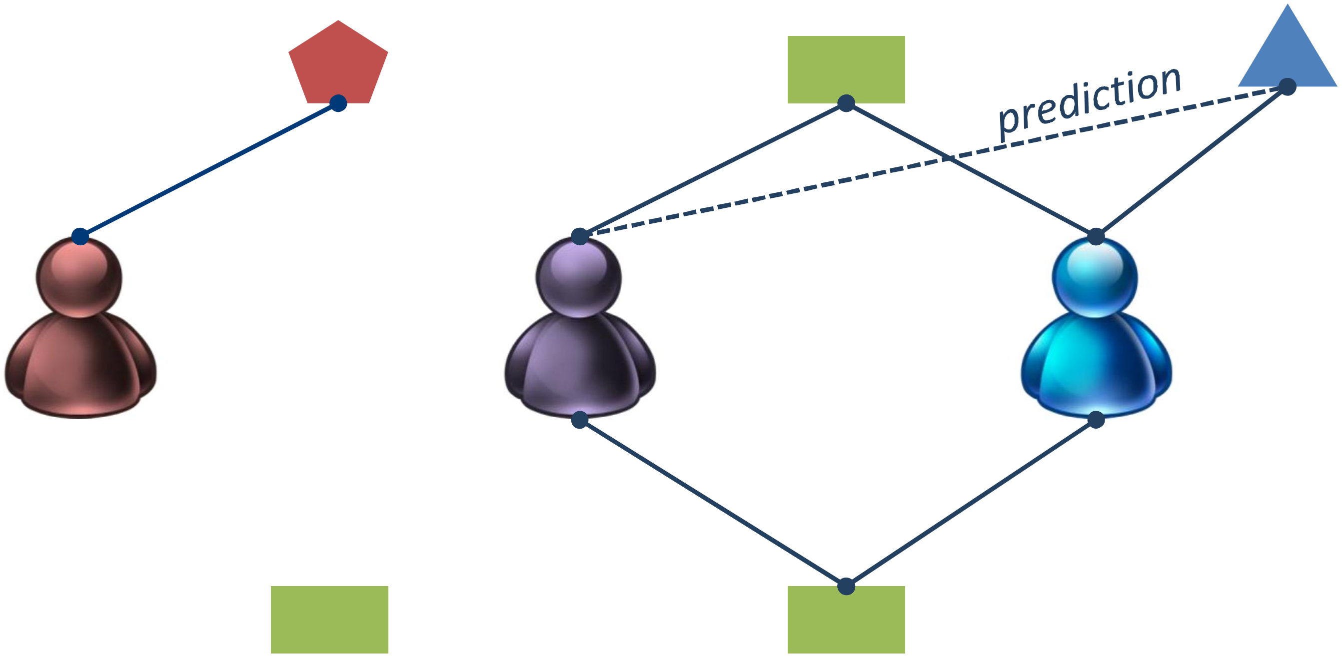


  - Collaborative filtering
  - Social filtering
  - …
  - Hybrid filtering

"You may like rock music if you like heavy metal"

Recommender Systems (3)

- Item suggestions can be obtained using several techniques:
  - Content-based
  - **Collaborative filtering**

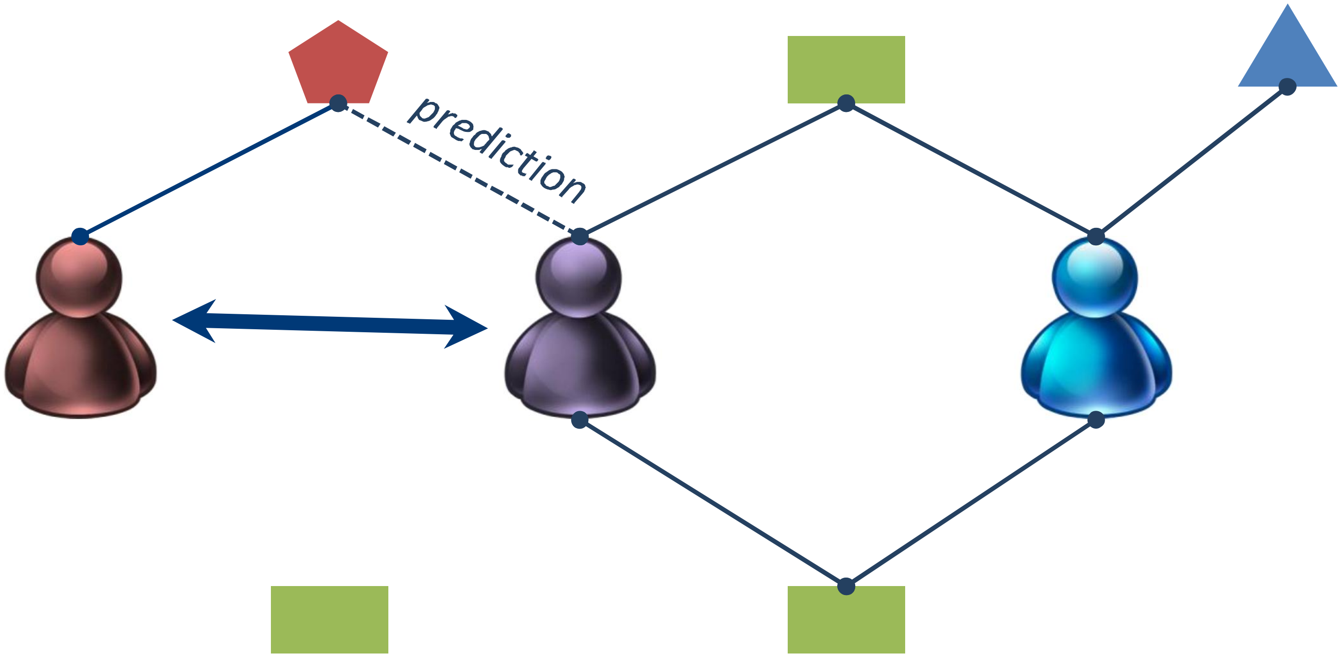


  - Social filtering
  - …
  - Hybrid filtering

"You may like classical music if you like heavy metal"

Recommender Systems (3)

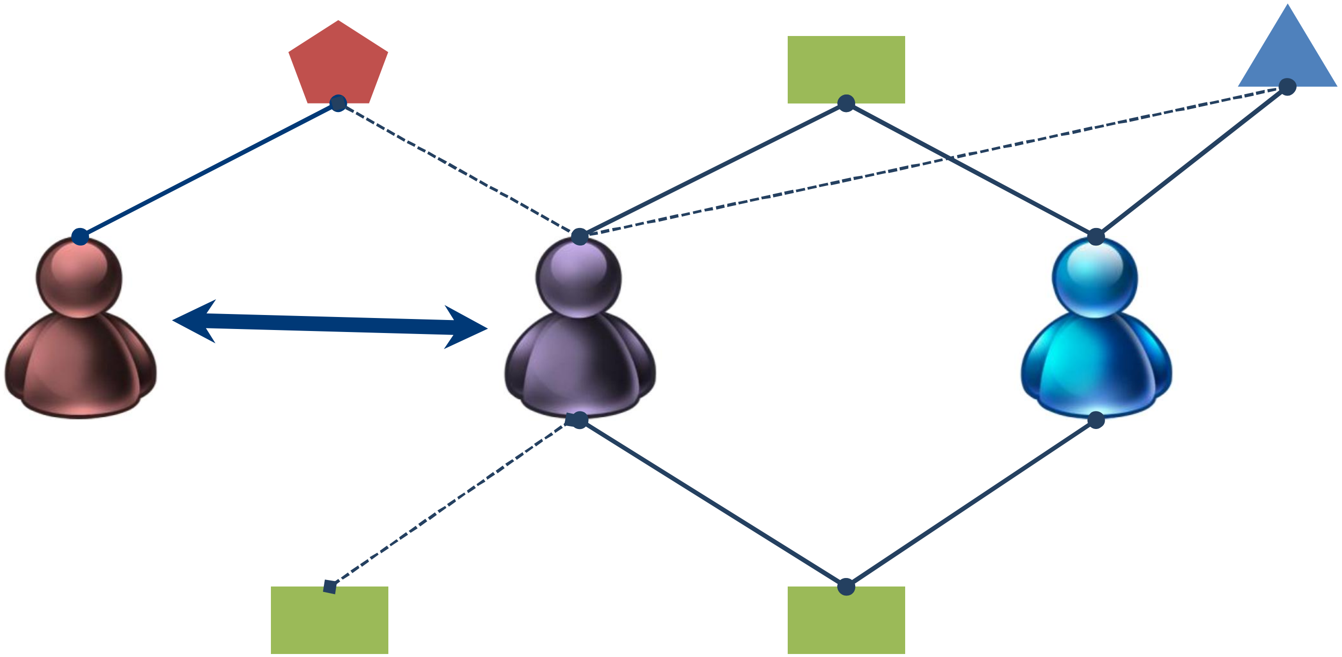
- Item suggestions can be obtained using several techniques:
  - Content-based
  - Collaborative filtering
  - **Social filtering**
  - …
  - Hybrid filtering



"You may like samba because your friend Marcelo likes it"

Recommender Systems (3)

- Item suggestions can be obtained using several techniques:
  - Content-based
  - Collaborative filtering
  - Social filtering
  - …
  - **Hybrid filtering**



Main research question

Is it possible to predict the performance of a specific recommendation approach or component?

- We need reliable measurements of performance
- We seek predictors with strong predictive power
- There are potential applications where these predictors may achieve an improvement in performance

Research goals

- RG1: Analysis and formalisation of how retrieval **performance** can be defined and evaluated in recommender systems
  - What is performance?
  - How should we measure performance?
- RG2: Adaptation and definition of **performance prediction** techniques to recommender systems
  - How can we estimate the performance of a recommender?
- RG3: **Application** of performance predictors to hybrid recommender systems
  - Where (and how) can we apply our performance predictors?

Proposal

- RG1: *Evaluating performance in recommender systems*
  - We analyse design alternatives in recommender evaluation and discuss differences with respect to IR
  - We detect resulting biases and propose designs to neutralise them
- RG2: *Predicting performance in recommender systems*
  - We show adaptations to recommendation of performance predictors from IR
  - We report strong predictive power between true and predicted performances
- RG3: *Applications*
  - We research applications of performance predictors to dynamic aggregations of information
  - We find that predictors with strong predictive power tend to obtain higher improvements in dynamic applications

Contents

- Part I – Evaluating performance in recommender systems
  - Performance evaluation in recommender systems
  - Experimental designs and biases
- Part II – Predicting performance in recommender systems
  - Performance prediction in Information Retrieval
  - Performance prediction in recommender systems
- Part III – Applications
  - Dynamic recommender ensembles
  - Neighbour selection and weighting in collaborative filtering
- Conclusions and future work

Contents

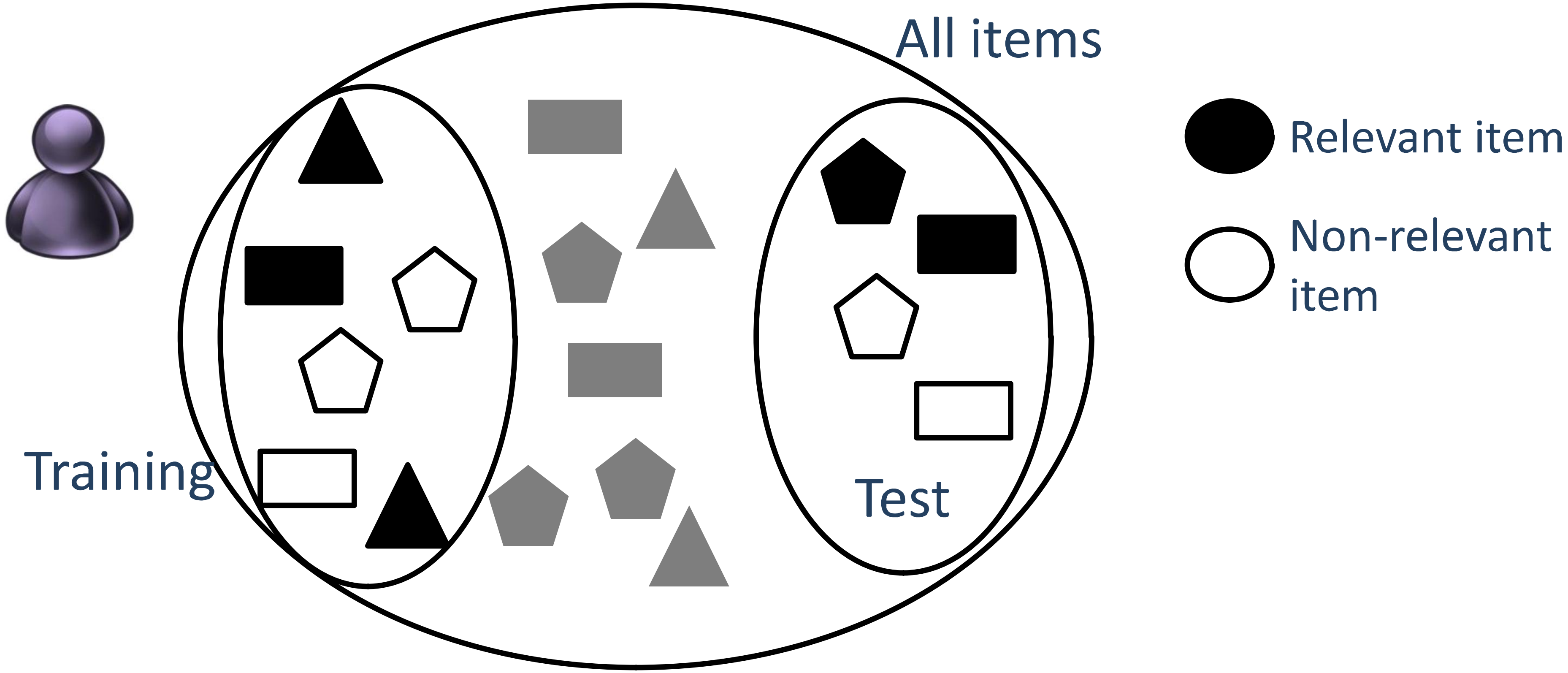
Performance evaluation in recommender systems (1)

- Error metrics have been dominant in the literature
  - Root Mean Square Error (RMSE), Mean Absolute Error (MAE)
- Now, ranking metrics are increasingly used
  - Precision, recall
- In general, a set of items are issued to the recommender and ranked according to the estimated preference
- Each experimental design would select a set of candidate items in different ways

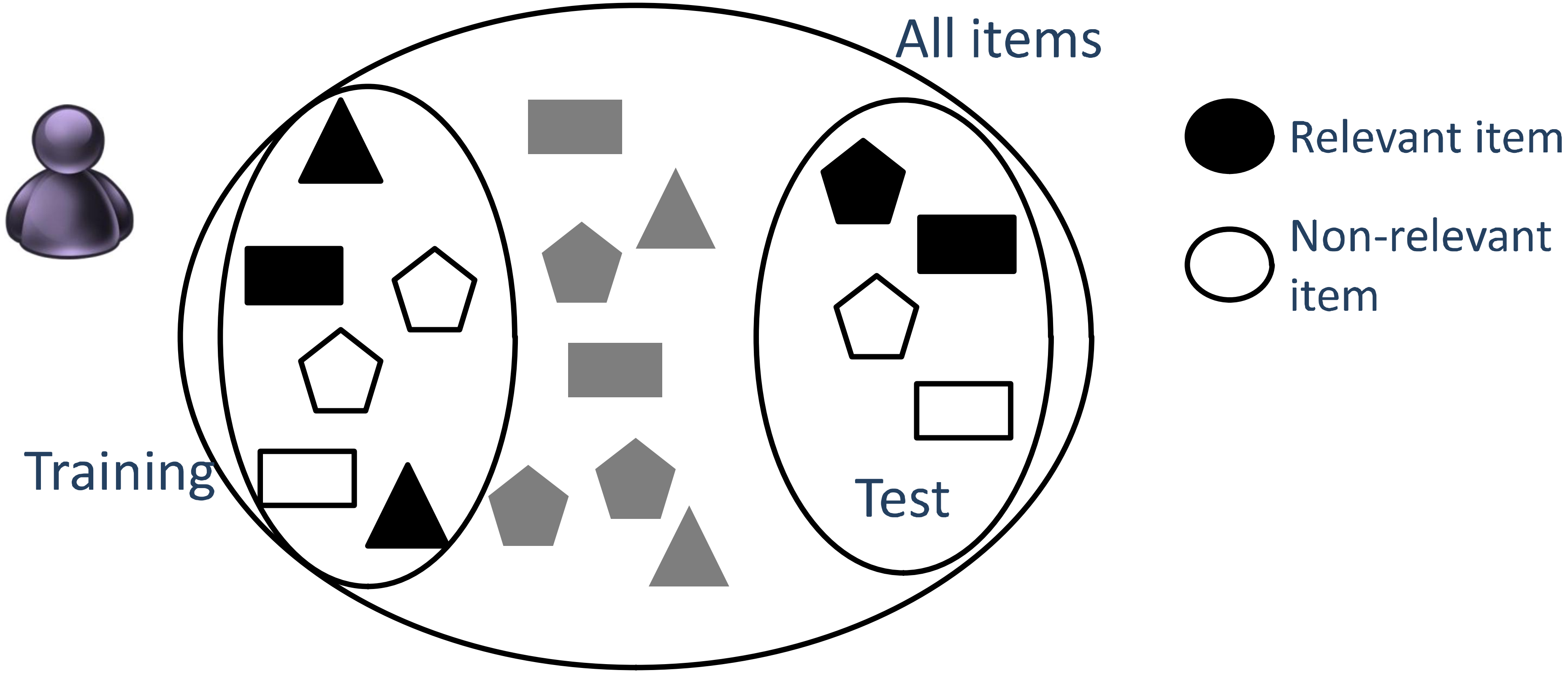
Experimental designs

- The adoption of IR methodologies is natural:
  - Query ≈ User
  - Document ≈ Item
  - Relevant ≈ Test (positive) rating
- However, there are differences in the evaluation settings:
  - The candidate answers
    - Retrieval: all the documents, the same for all the queries
    - Recommendation: training/test split, a target item set different for each user
  - Relevance / ground truth
    - Retrieval: assumed to be reasonably complete, objective
    - Recommendation: highly incomplete, subjective

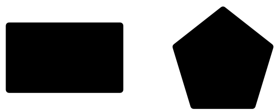
Candidate item selection (1)



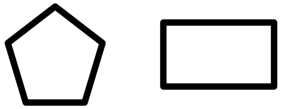
Candidate item selection (2)



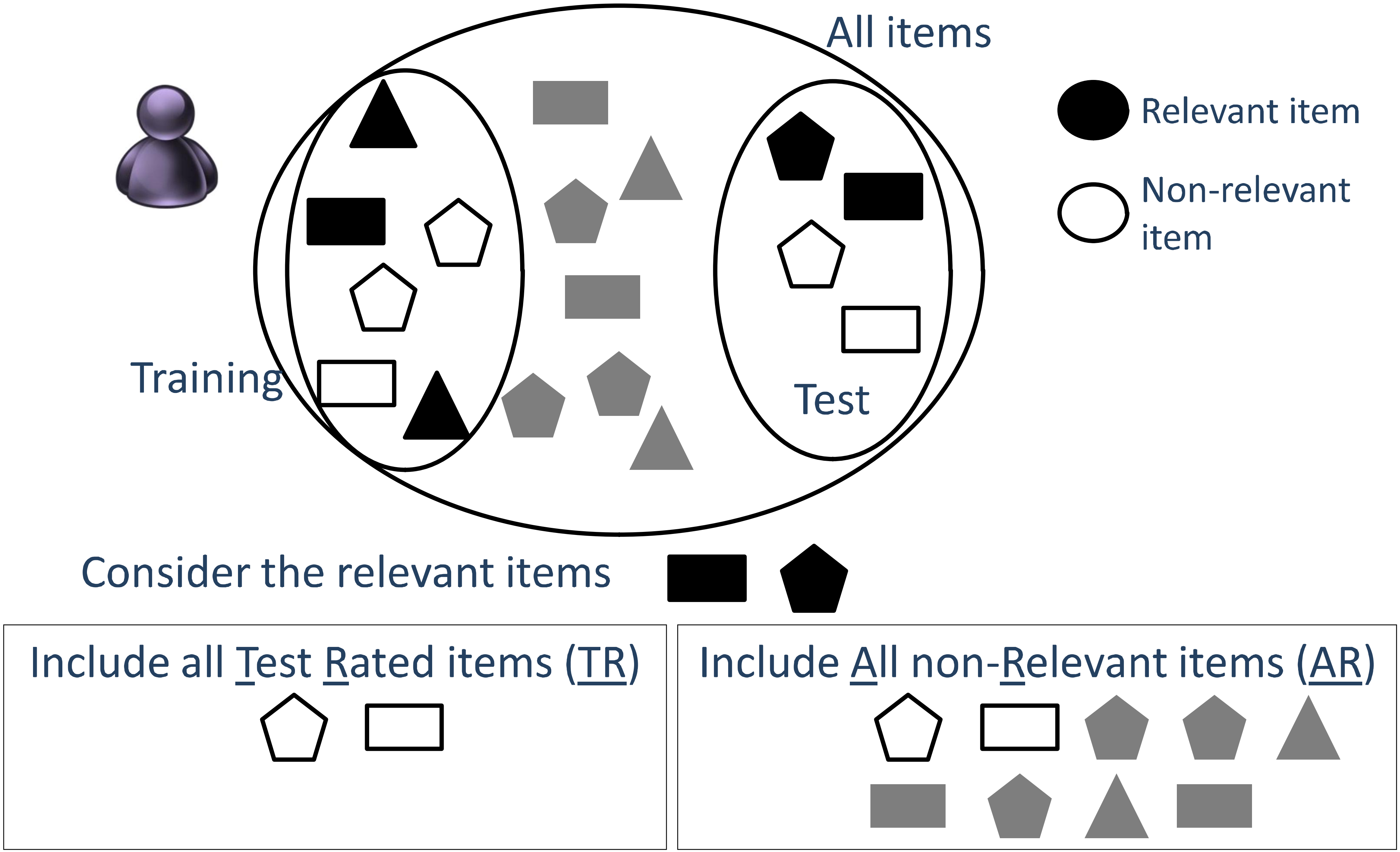
Consider the relevant items



Include all Test Rated items (TR)



Candidate item selection (3)



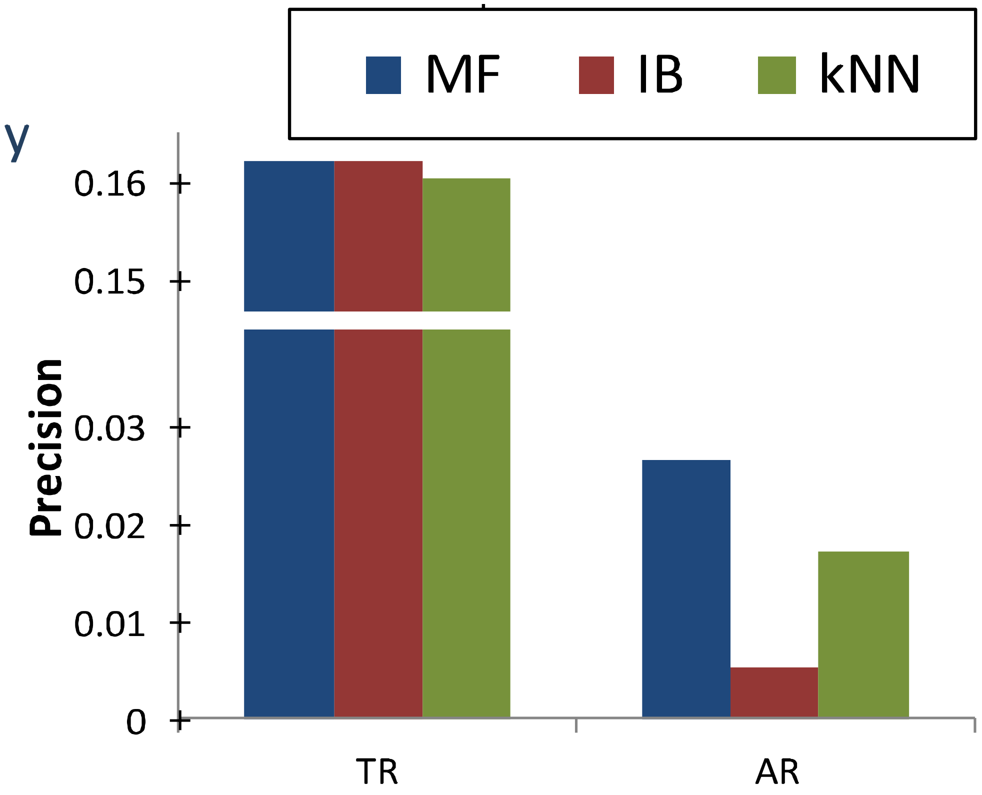
Could the candidate item selection affect the measured performance of the system?

Results with different candidate item strategies

- In the literature

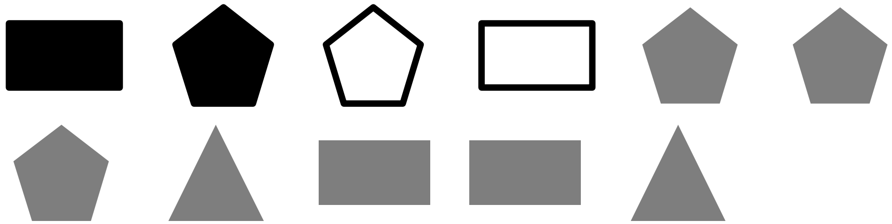
Different results are reported depending on the selected items to rank

- We have compared the TR and AR designs
  - Different <u>absolute</u> values
  - Recommenders <u>compare</u> differently



Experimental designs

- We discard TR because it highly overestimates precision
- In this thesis, we use the following designs (methodologies):
  - All non-Relevant and All Relevant test items: **AR**



  - One Relevant test item per ranking: **1R**. Plus a fixed number of non-relevant items

(Cremonesi et al., 2010)

Experimental designs and biases

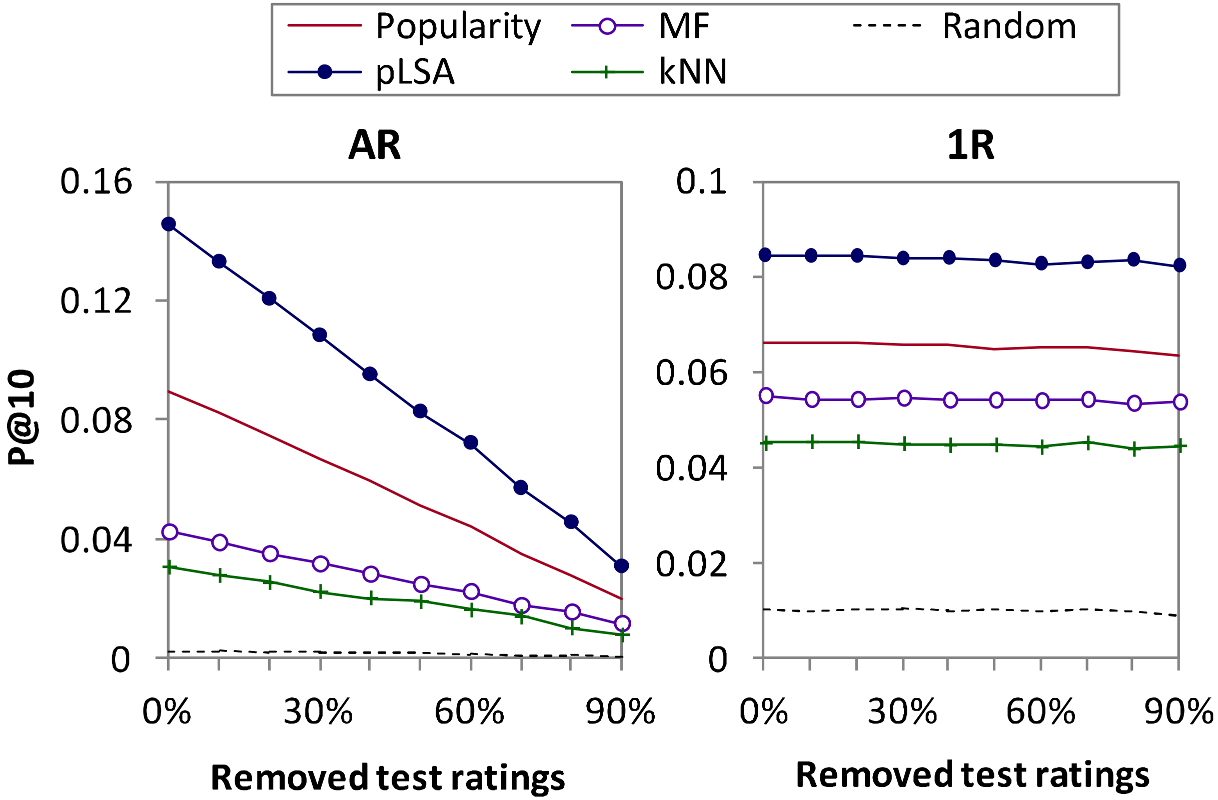
- We have identified the following biases in the AR and 1R designs:
  - **Sparsity bias**: metric values change depending on the ratio of relevant items
  - **Popularity bias**: metrics favour the overall satisfaction of the users

- We study the effect of these biases
  - Analytically (in terms of expected precision)
  - Empirically

- Experimental settings
  - Dataset: MovieLens, Last.fm
  - Evaluation metric: Precision at 10
  - Recommenders: personalised (kNN, MF, pLSA) and non-personalised (Popularity, Random)

Sparsity bias

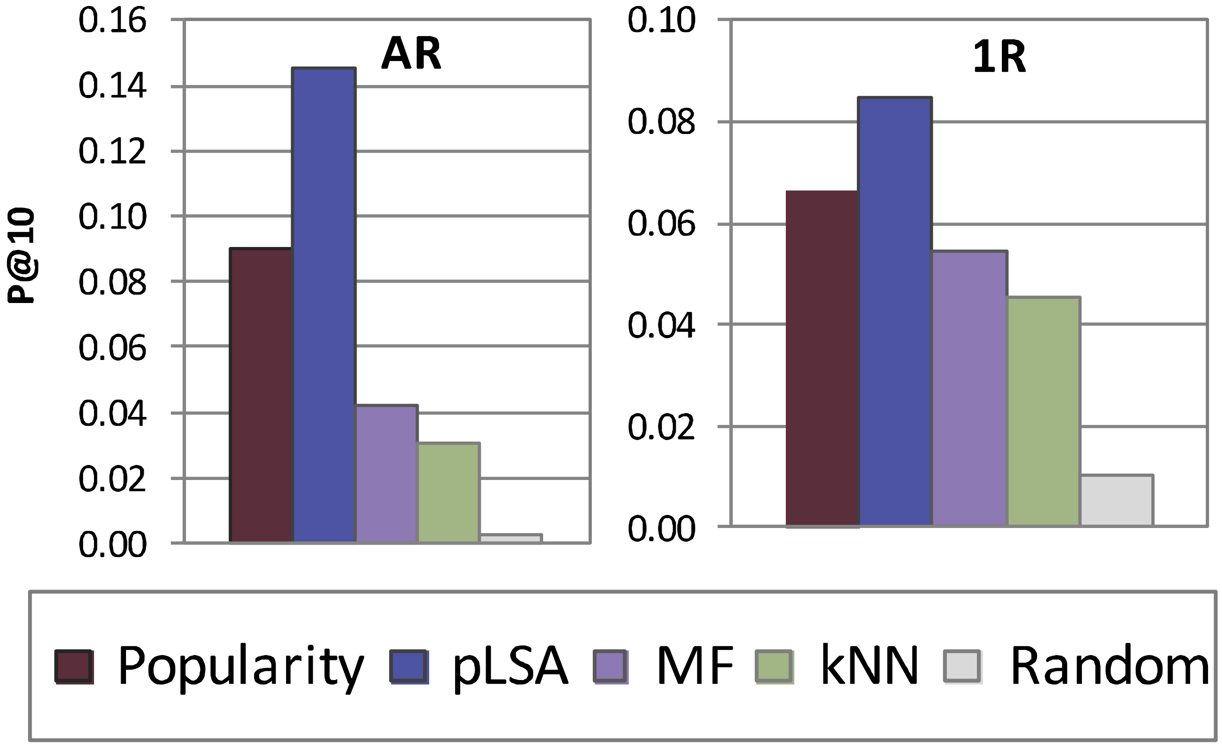
- Experiments
  - Change the density of known relevance



- Conclusions
  - Precision values in AR are useful only for comparative purposes
  - Precision values in 1R are not sensitive to the sparsity level

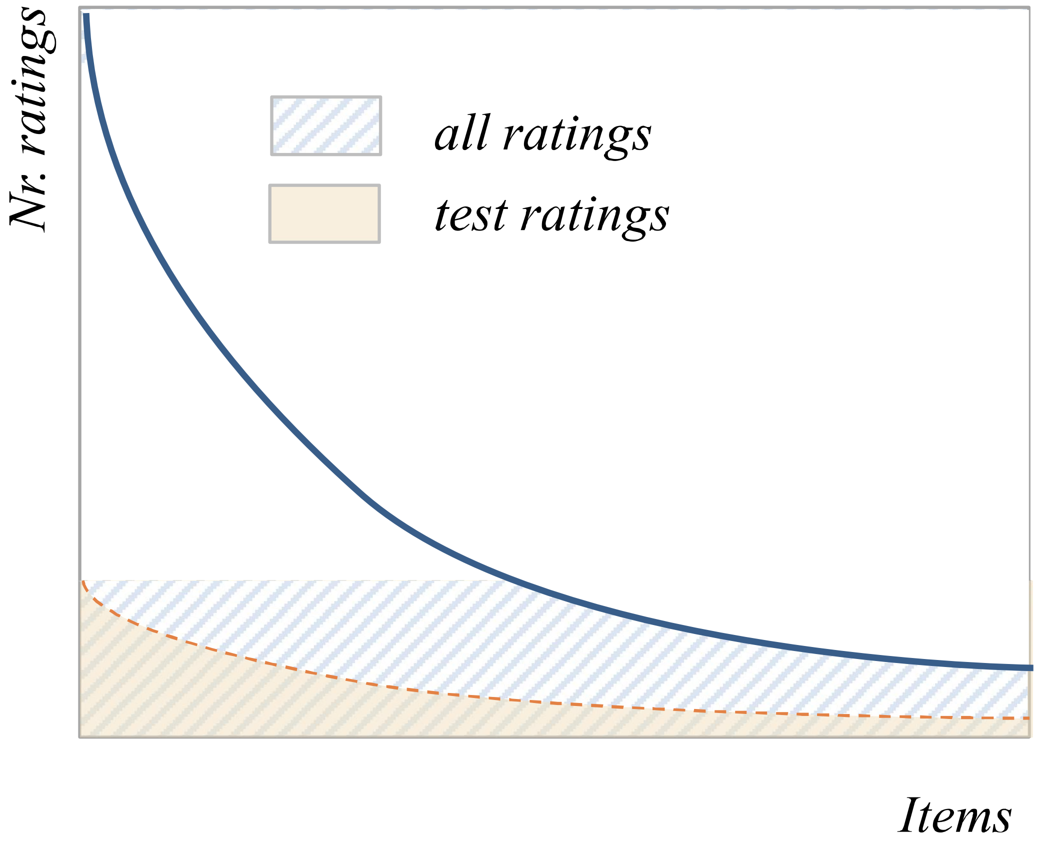
Popularity bias (1)

- The popularity-based recommender outperforms other techniques
- Empirical evidence
  - Both methodologies are sensitive to the effect of popularity



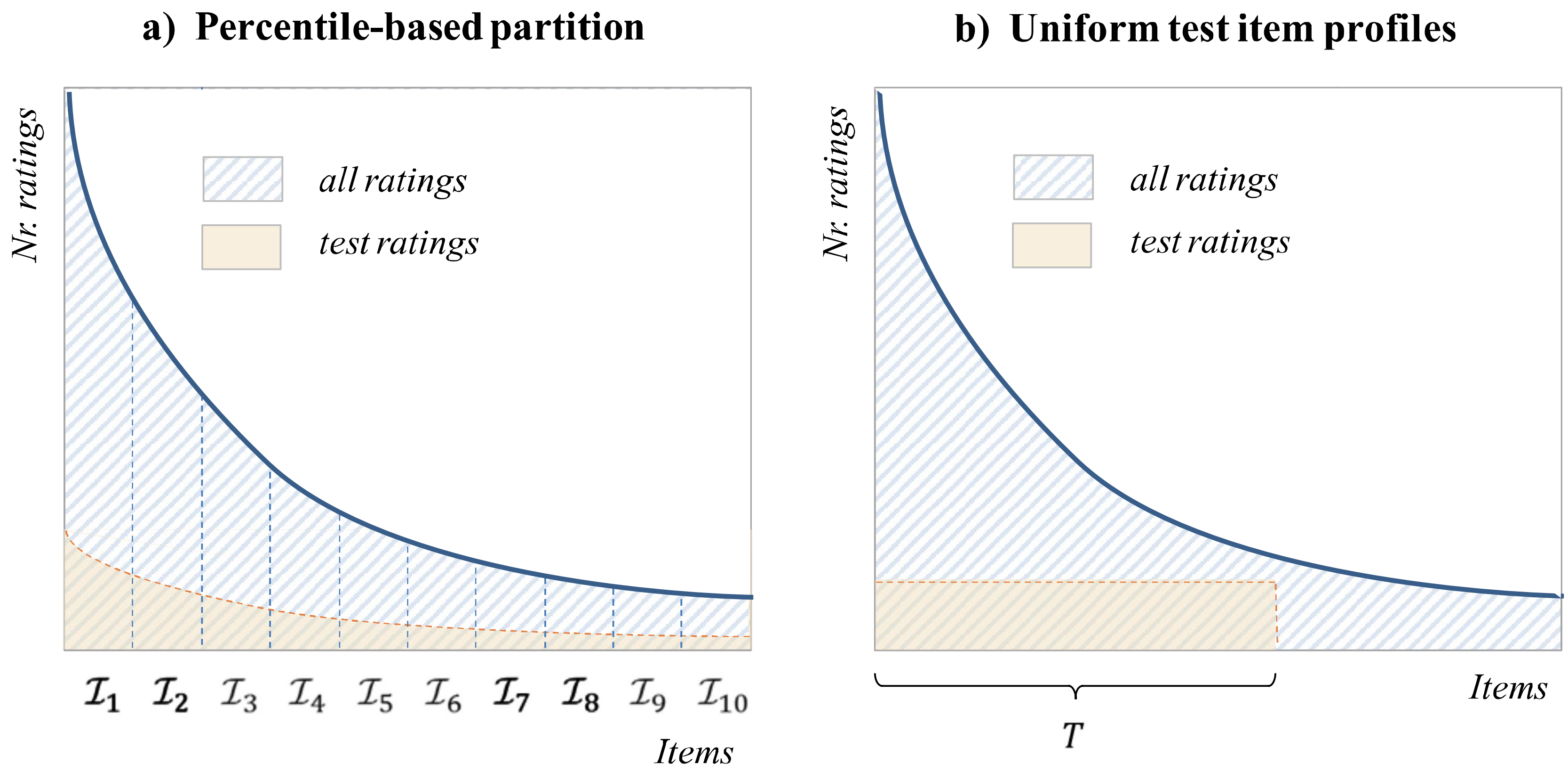
Popularity bias (2)

- The popularity-based recommender outperforms other techniques
  - Due to statistical reasons, popular items appear more often in the test set
  - Average precision metrics tend to favour the satisfaction of majorities



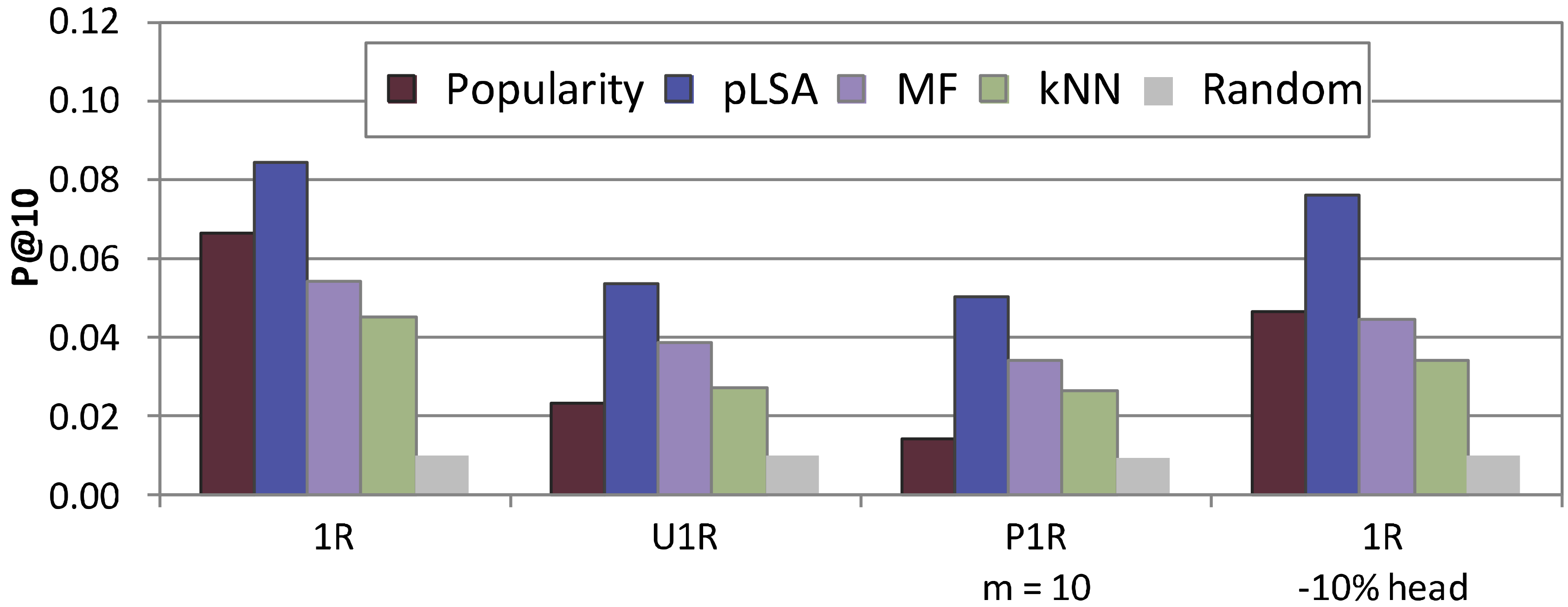
Overcoming the popularity bias

- We propose two methodologies to overcome the popularity bias
  - **Percentile-based partition** (P1R): the items are grouped according to their popularity
  - **Uniform test item profiles** (U1R): all the items have the same amount of test ratings



Experiments

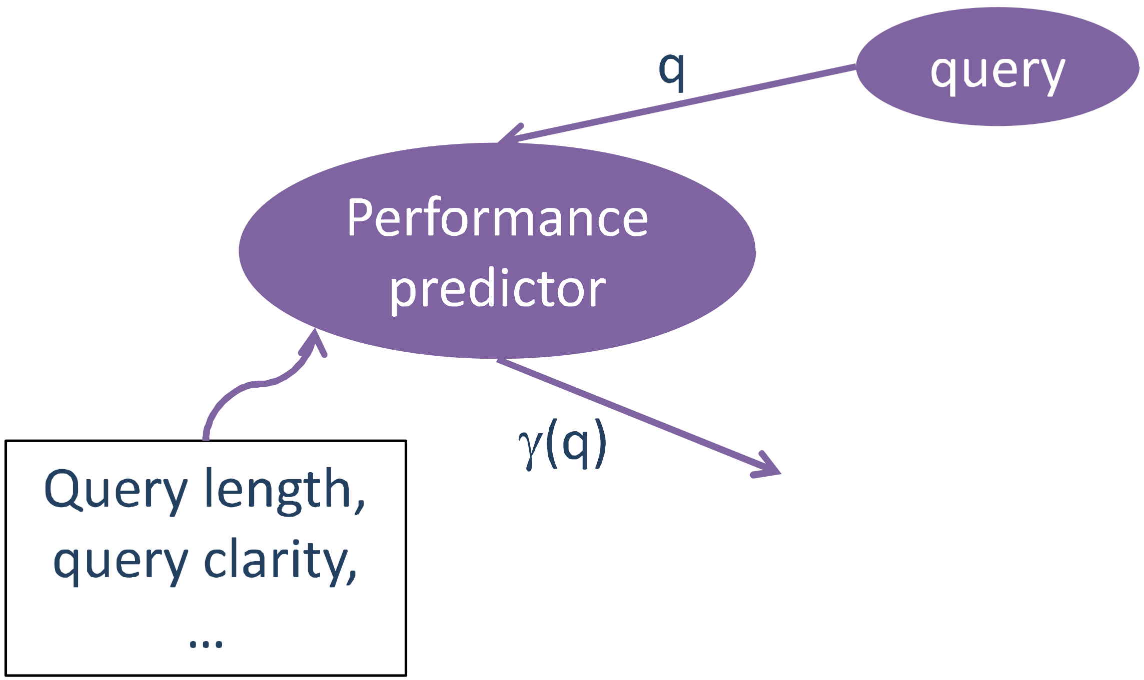
- Comparison of results: biased vs. unbiased experimental designs



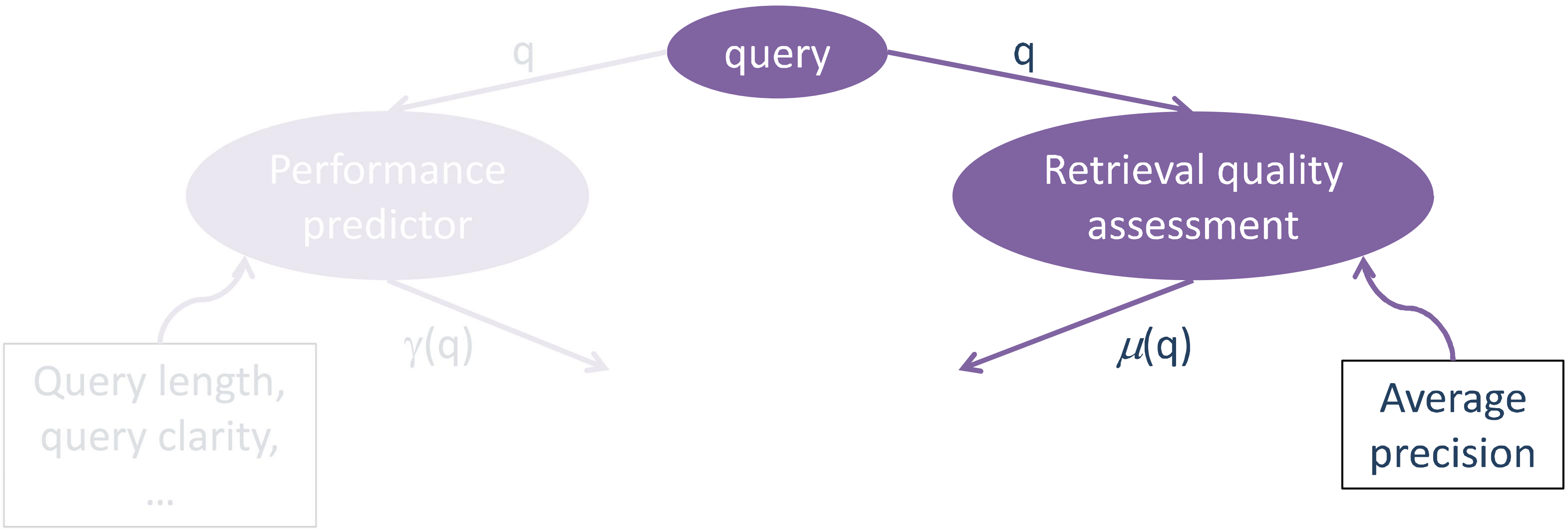
- Conclusions
  - U1R and P1R discriminate between pure popularity-based and personalised recommendation
  - Better discrimination than removing the 10% of most popular items from test (Cremonesi et al., 2010)

Contents

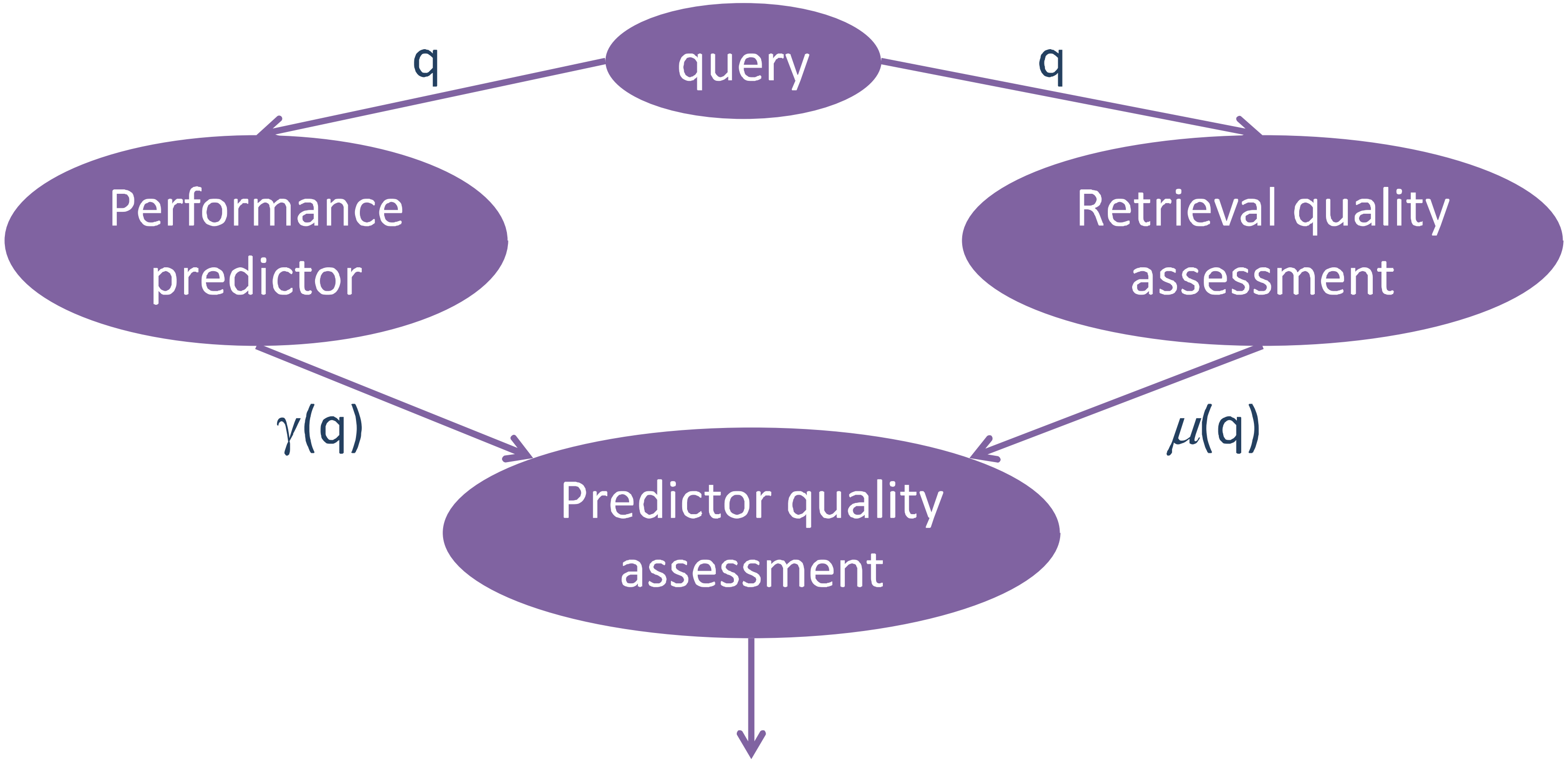
Performance prediction in Information Retrieval (1)



Performance prediction in Information Retrieval (2)



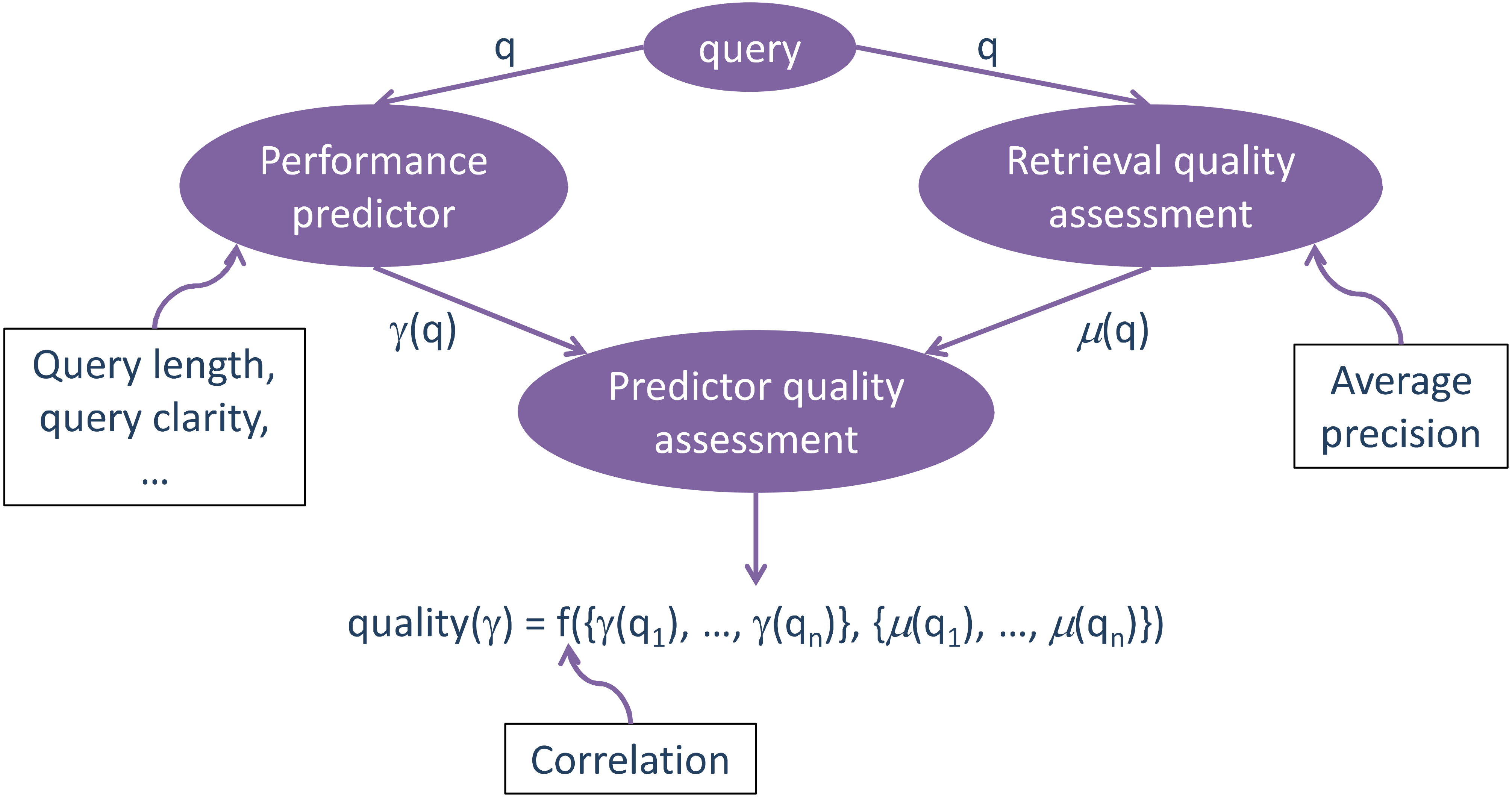
Performance prediction in Information Retrieval (3)



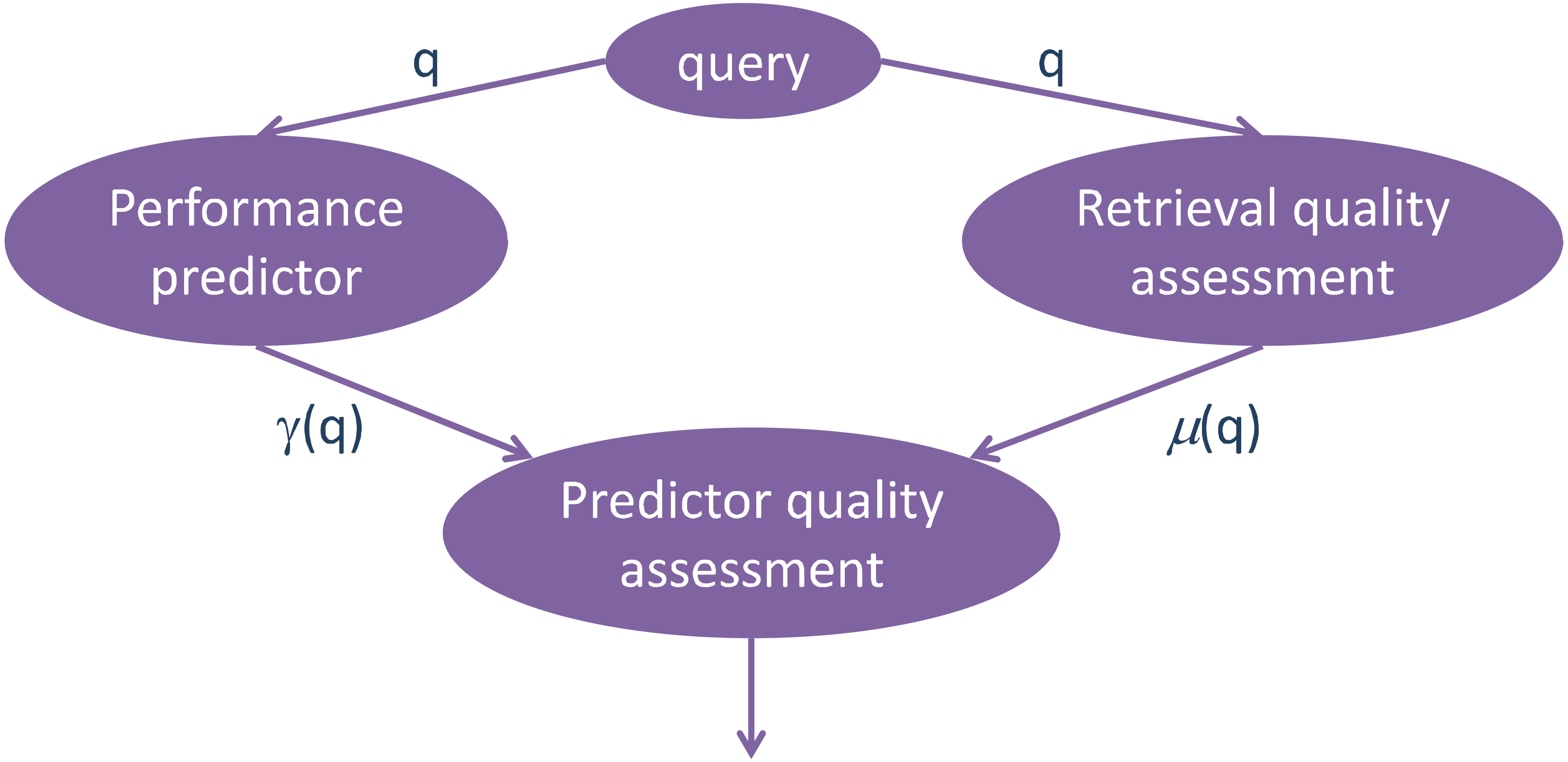
$$\text{quality}(\gamma) = f(\{\gamma(q_1), \ldots, \gamma(q_n)\}, \{\mu(q_1), \ldots, \mu(q_n)\})$$

Correlation

Performance prediction in Information Retrieval (4)



Performance prediction in Information Retrieval (5)



$$\text{quality}(\gamma) = f(\{\gamma(q_1), \ldots, \gamma(q_n)\}, \{\mu(q_1), \ldots, \mu(q_n)\})$$

- Some applications
  - Query expansion: deciding which queries should be expanded
  - Query rephrasing: providing feedback to the user
  - Rank aggregation: combining results from different retrieval models

A performance predictor in IR

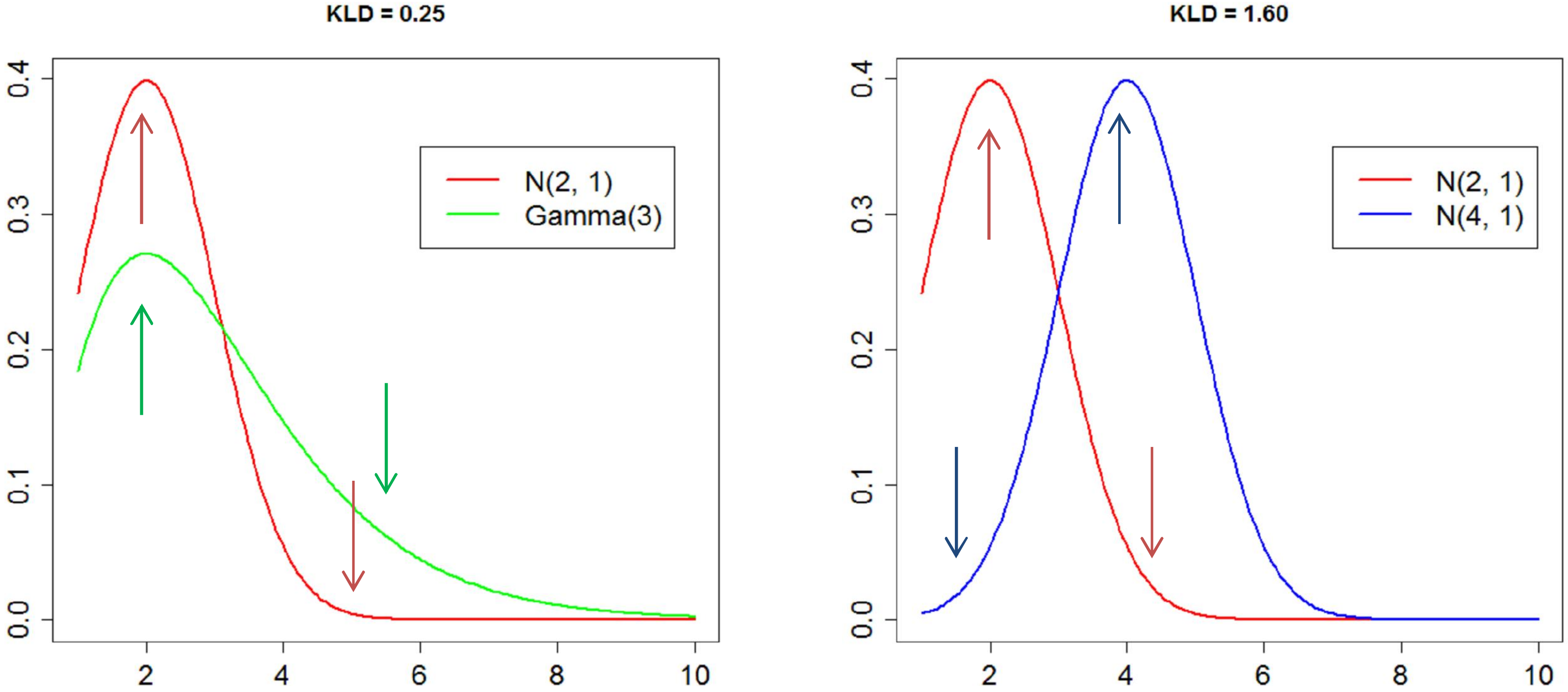
Query clarity

Query clarity

- It measures the (Kullback-Leibler) divergence between the query and the collection language model

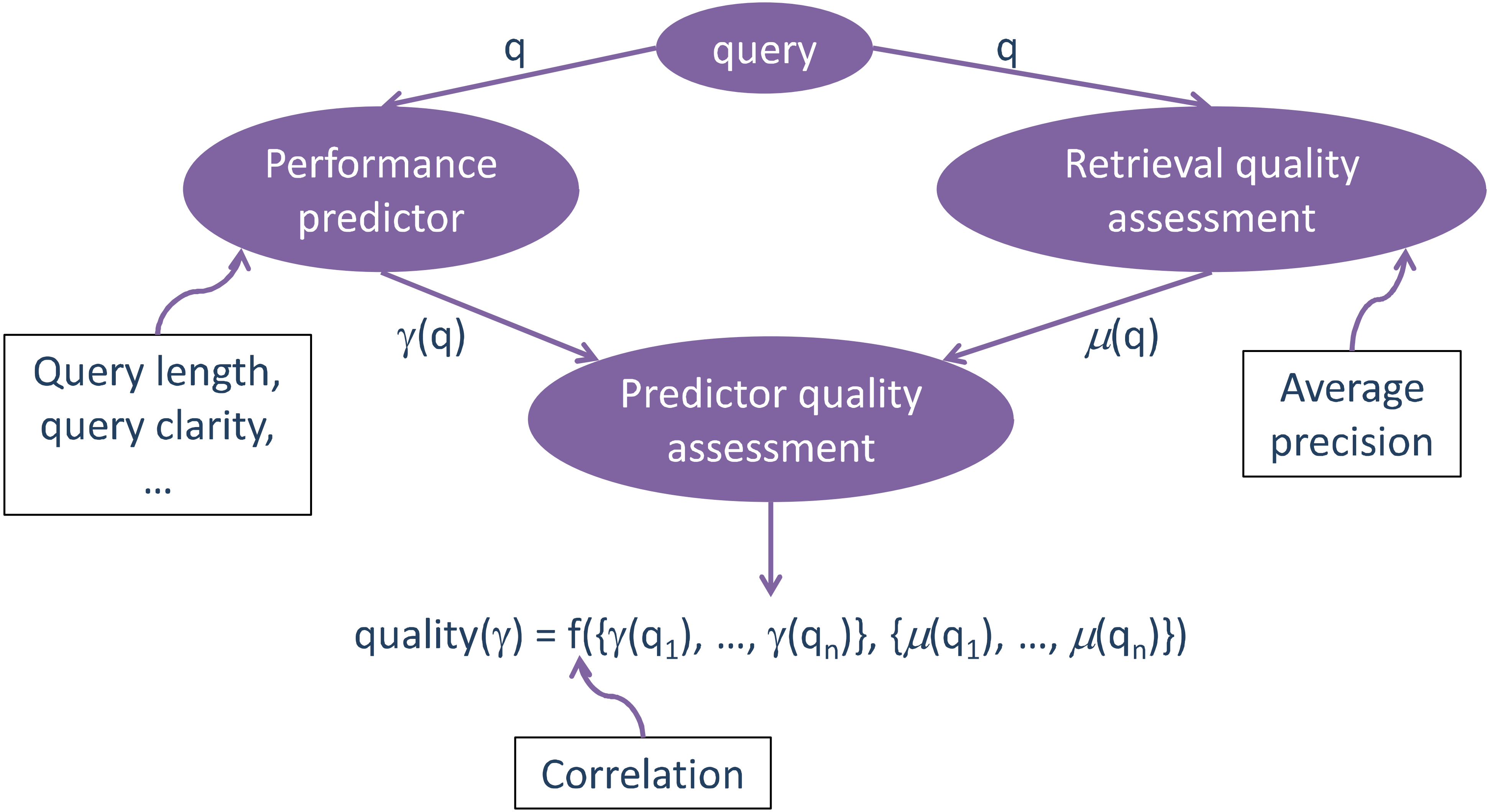
$$\text{clarity}(q) = \sum_{w \in V} p(w \mid q) \log \frac{p(w \mid q)}{p(w)}$$

- Clear queries are those whose distributions are different from the collection's distribution

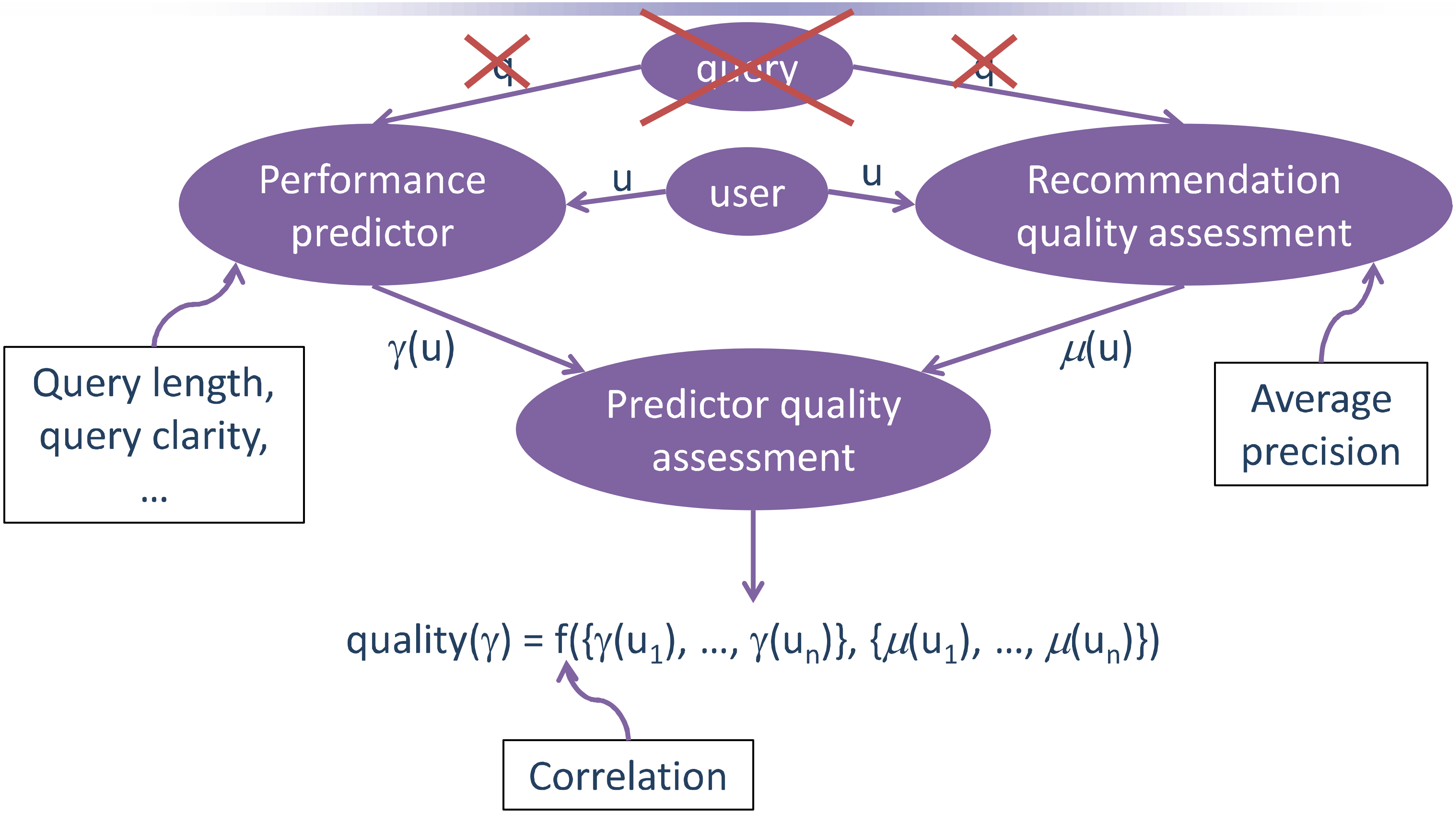


Performance prediction in recommender systems

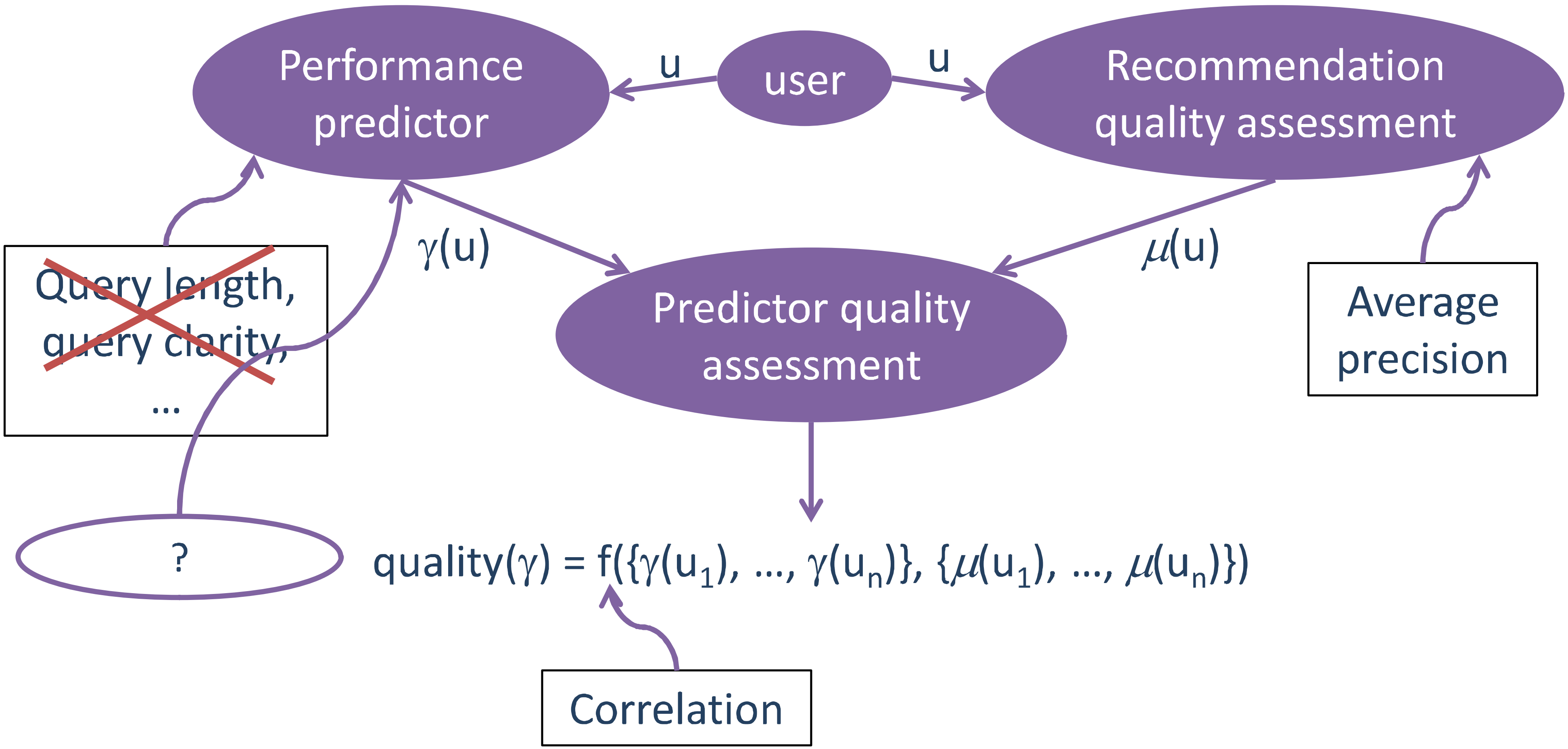
Performance prediction in Information Retrieval



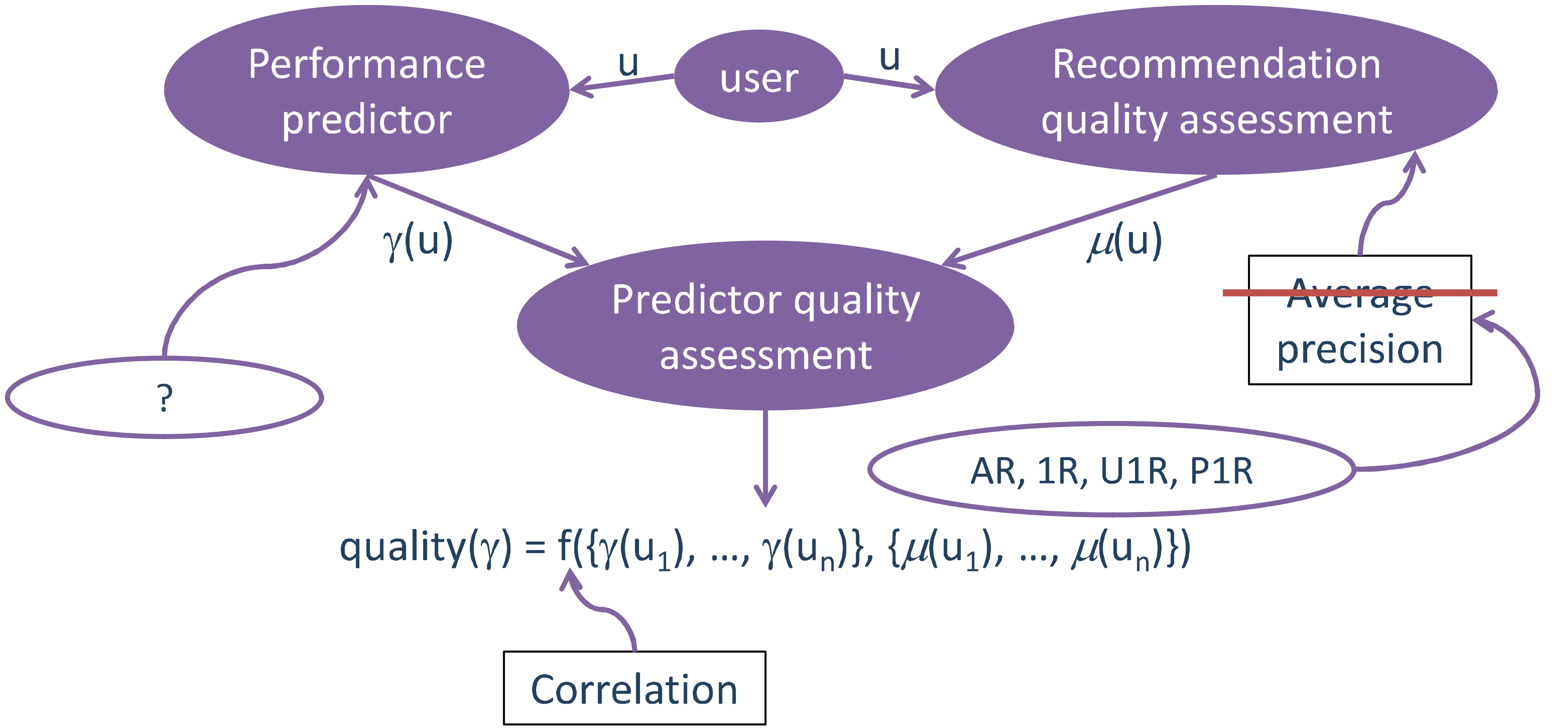
Performance prediction in recommender systems (1)



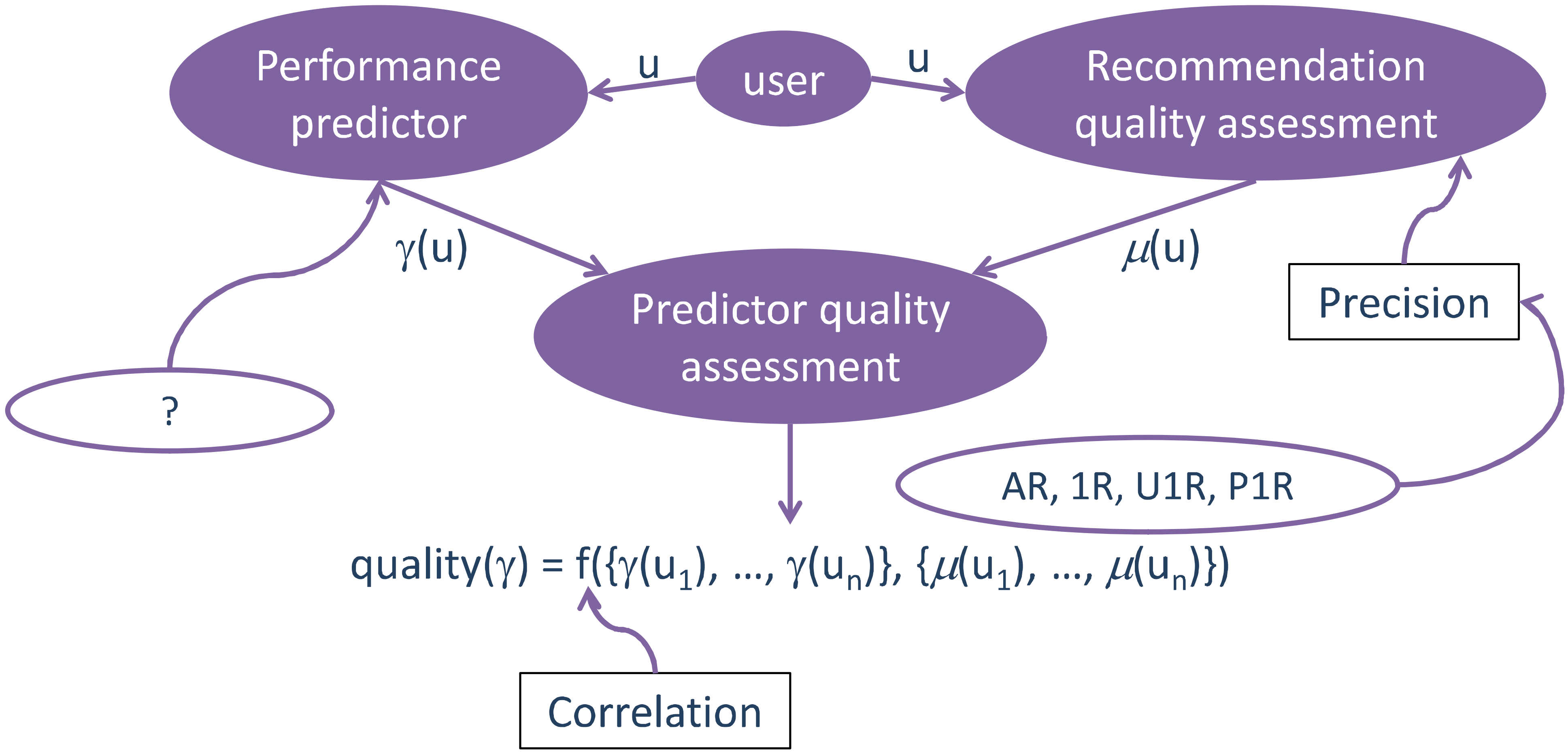
Performance prediction in recommender systems (2)



Performance prediction in recommender systems (3)

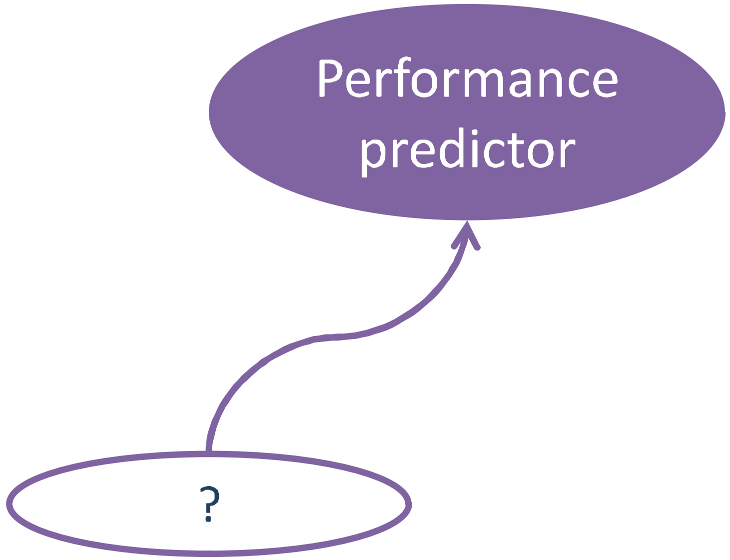


Performance prediction in recommender systems (4)



Performance prediction in recommender systems (5)

- We propose definitions of user predictors
  - Based on rating data
  - Based on log data
  - Based on social data
- We use
  - Query clarity adaptations
  - Measures from Information Theory (e.g., entropy)
  - Social graph metrics (e.g., PageRank, HITS, centrality)



User clarity (1)

- Query clarity

$$\text{clarity}(q) = \sum_{w \in V} p(w \mid q) \log \frac{p(w \mid q)}{p(w)}$$

- User clarity

$$\text{clarity}(u) = \sum_{x \in X} p(x \mid u) \log \frac{p(x \mid u)}{p(x)}$$

  - Freedom to select the vocabulary space X

User clarity (2)

- Query clarity

$$\text{clarity}(q) = \sum_{w \in V} p(w \mid q) \log \frac{p(w \mid q)}{p(w)}$$

- <u>Generalized</u> user clarity

$$\text{clarity}(u) = \mathrm{E}_{\theta}\left[\sum_{x \in X} p(x \mid u, \theta) \log \frac{p(x \mid u, \theta)}{p(x \mid \theta)}\right]$$

  - Freedom to select the vocabulary space X
  - Possibility to introduce a context variable θ in some formulations
  - They let <u>capture different aspects</u> of the user

User clarity for rating data

- User clarity

$$\text{clarity}(u) = \mathrm{E}_{\theta}\left[\sum_{x \in X} p(x \mid u, \theta) \log \frac{p(x \mid u, \theta)}{p(x \mid \theta)}\right]$$

Rating data: (user, item, rating)

Rating based

$$\sum_{r} p(r \mid u) \log \frac{p(r \mid u)}{p(r)}$$

Item based

$$\sum_{i} p(i \mid u) \log \frac{p(i \mid u)}{p(i)}$$

Item-and-rating based

$$\sum_{r,i} p(i)\, p(r \mid u, i) \log \frac{p(r \mid u, i)}{p(r \mid i)}$$

User clarity for rating data

- User clarity

$$\text{clarity}(u) = \mathrm{E}_{\theta}\left[\sum_{x \in X} p(x \mid u, \theta) \log \frac{p(x \mid u, \theta)}{p(x \mid \theta)}\right]$$

Rating data: (user, item, rating)

Rating based

$$\sum_{r} p(r \mid u) \log \frac{p(r \mid u)}{p(r)}$$

Item based

$$\sum_{i} p(i \mid u) \log \frac{p(i \mid u)}{p(i)}$$

Item-and-rating based

$$\sum_{r,i} p(i)\, p(r \mid u, i) \log \frac{p(r \mid u, i)}{p(r \mid i)}$$

User clarity for log data

- User clarity

$$\text{clarity}(u) = \mathrm{E}_{\theta}\left[\sum_{x \in X} p(x \mid u, \theta) \log \frac{p(x \mid u, \theta)}{p(x \mid \theta)}\right]$$

Log data: (user, item, timestamp)

Frequency based

$$\sum_{i} p(i \mid u) \log \frac{p(i \mid u)}{p(i)}$$

Item space in user clarity

Item based

$$\sum_i p(i \mid u) \log \frac{p(i \mid u)}{p(i)}$$

$$p(i \mid u) = \sum_r p(i \mid u, r)\, p(r \mid u)$$

Frequency based

$$\sum_i p(i \mid u) \log \frac{p(i \mid u)}{p(i)}$$

$$p(i \mid u) = \frac{freq(i, u)}{\sum_{j \in I_u} freq(j, u)}$$

Temporal dimension for user clarity

- User clarity

$$\text{clarity}(u) = \mathrm{E}_{\theta}\left[\sum_{x \in X} p(x \mid u, \theta) \log \frac{p(x \mid u, \theta)}{p(x \mid \theta)}\right]$$

Log data: (user, item, timestamp)

Time based

$$\sum_{t} p(t \mid u) \log \frac{p(t \mid u)}{p(t)}$$

Item-and-time based

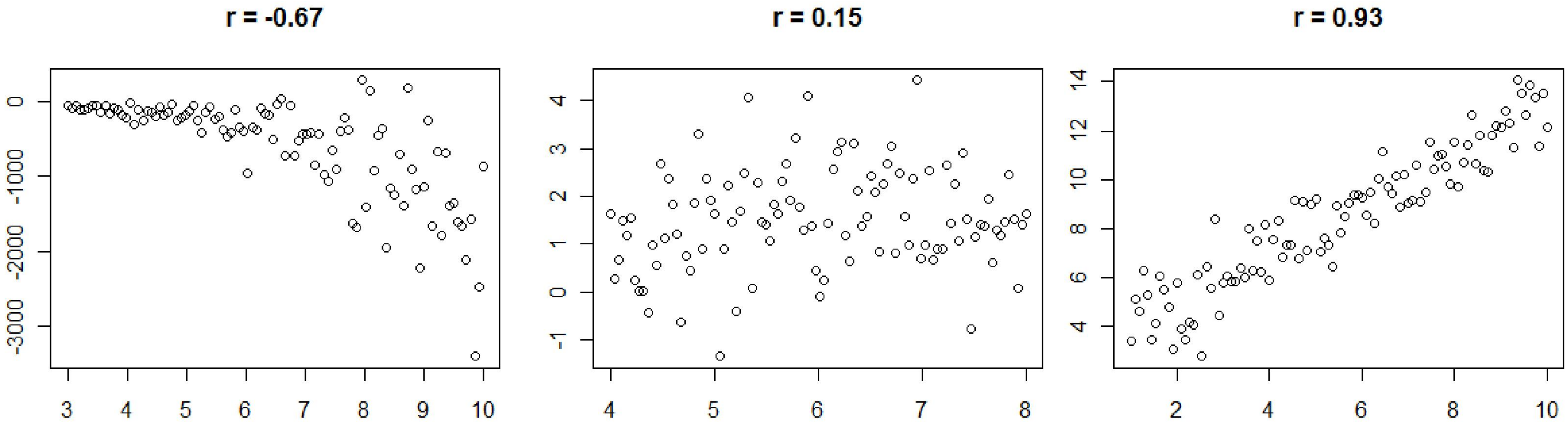
$$\sum_{t,i} p(i)\, p(t \mid u, i) \log \frac{p(t \mid u, i)}{p(t \mid i)}$$

User clarity

What is the predictive power of these models?

Experiments

- The predictive power is measured by the correlation with a metric of actual performance
- Experimental configuration
  - Performance metric: Precision at 10
  - Correlation coefficient: Pearson's r



Experiments

- The predictive power is measured by the correlation with a metric of actual performance

- Experimental configuration
  - Performance metric: Precision at 10
  - Correlation coefficient: Pearson's r
  - Evaluation methodologies: AR, 1R, U1R, P1R

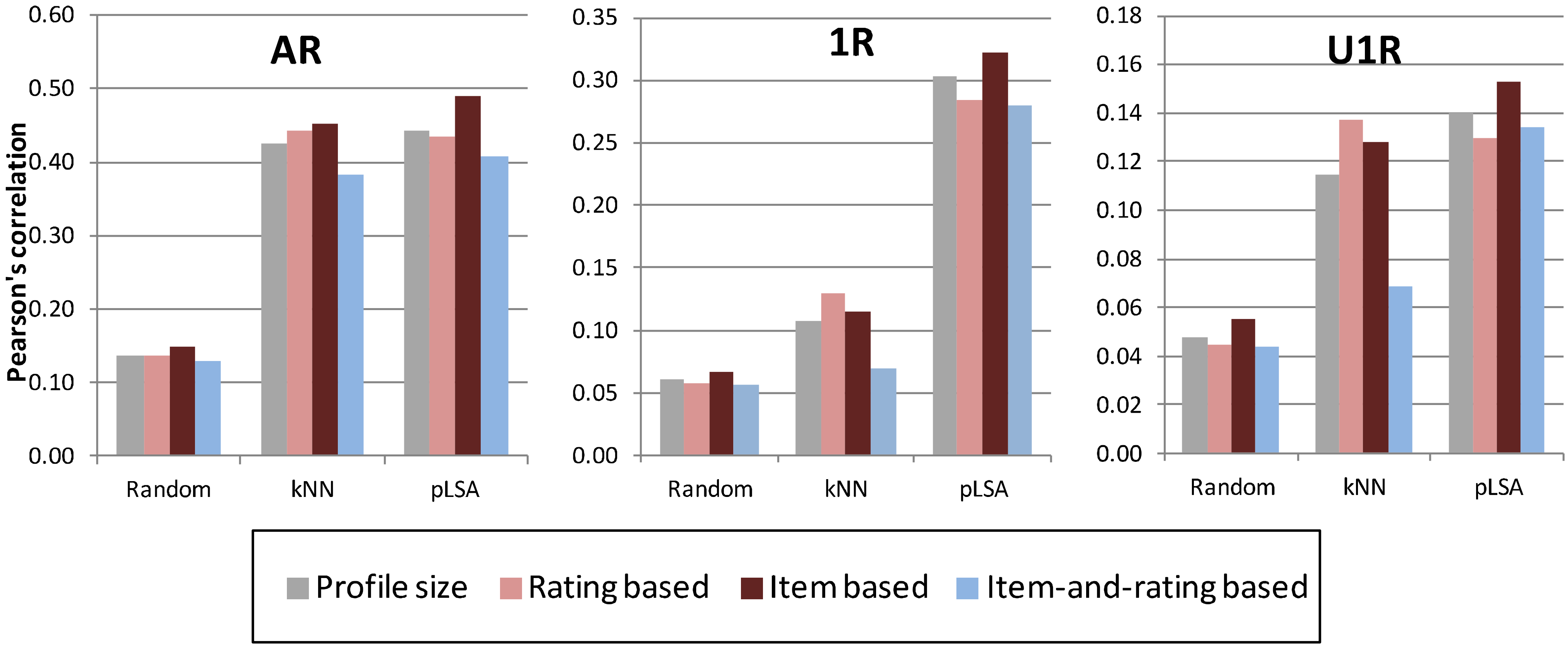
Are the proposed predictors sensitive to the statistical biases detected in some of these methodologies?

  - Datasets: MovieLens (ratings), Last.fm (logs), CAMRa (social)

Are the proposed predictors equally effective depending on the type of data?

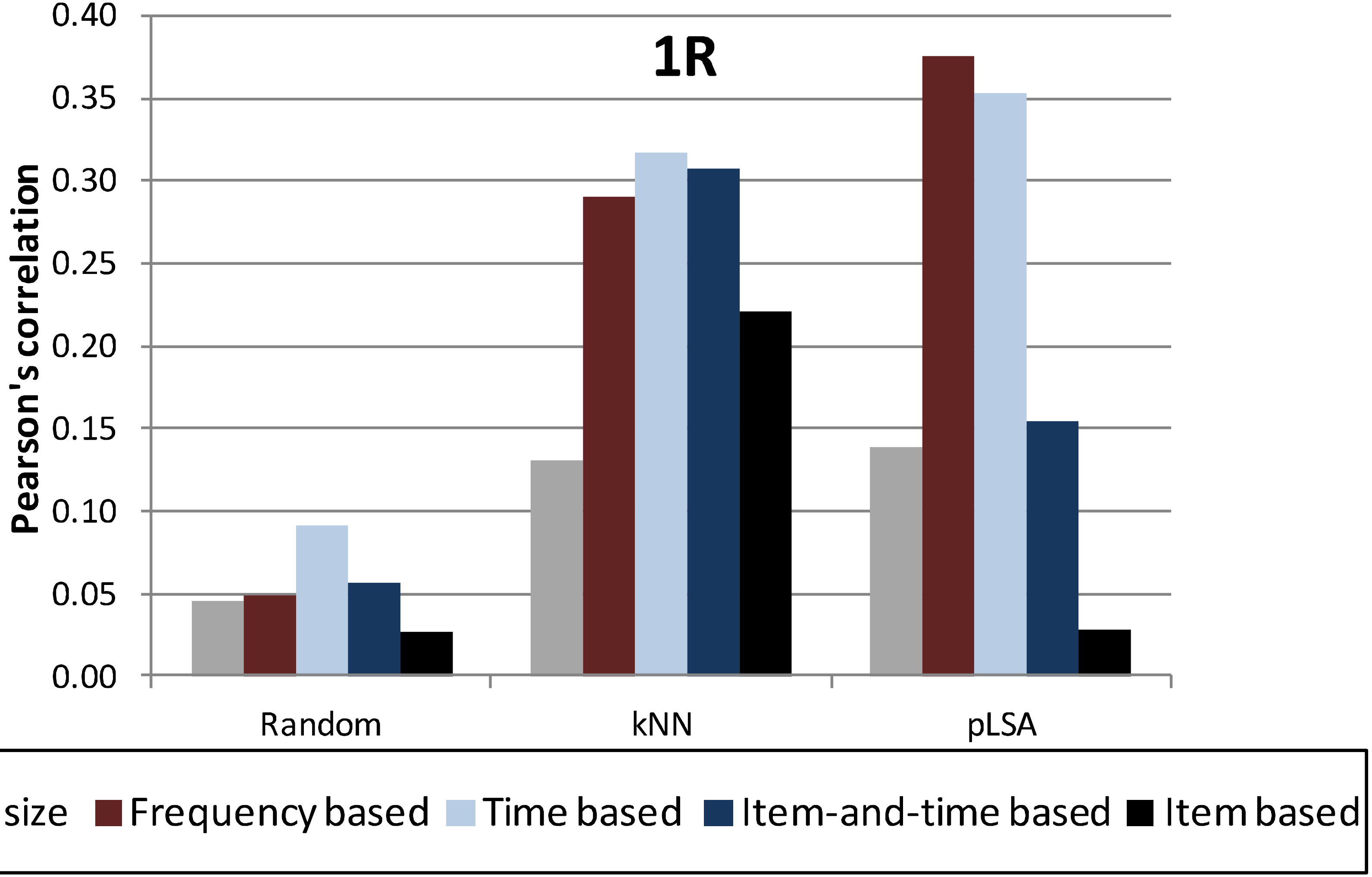
Experiments with rating data

- **User clarity** predictors
  - are particularly effective for rating data
  - achieve good results with unbiased experimental designs (similar with the P1R design)



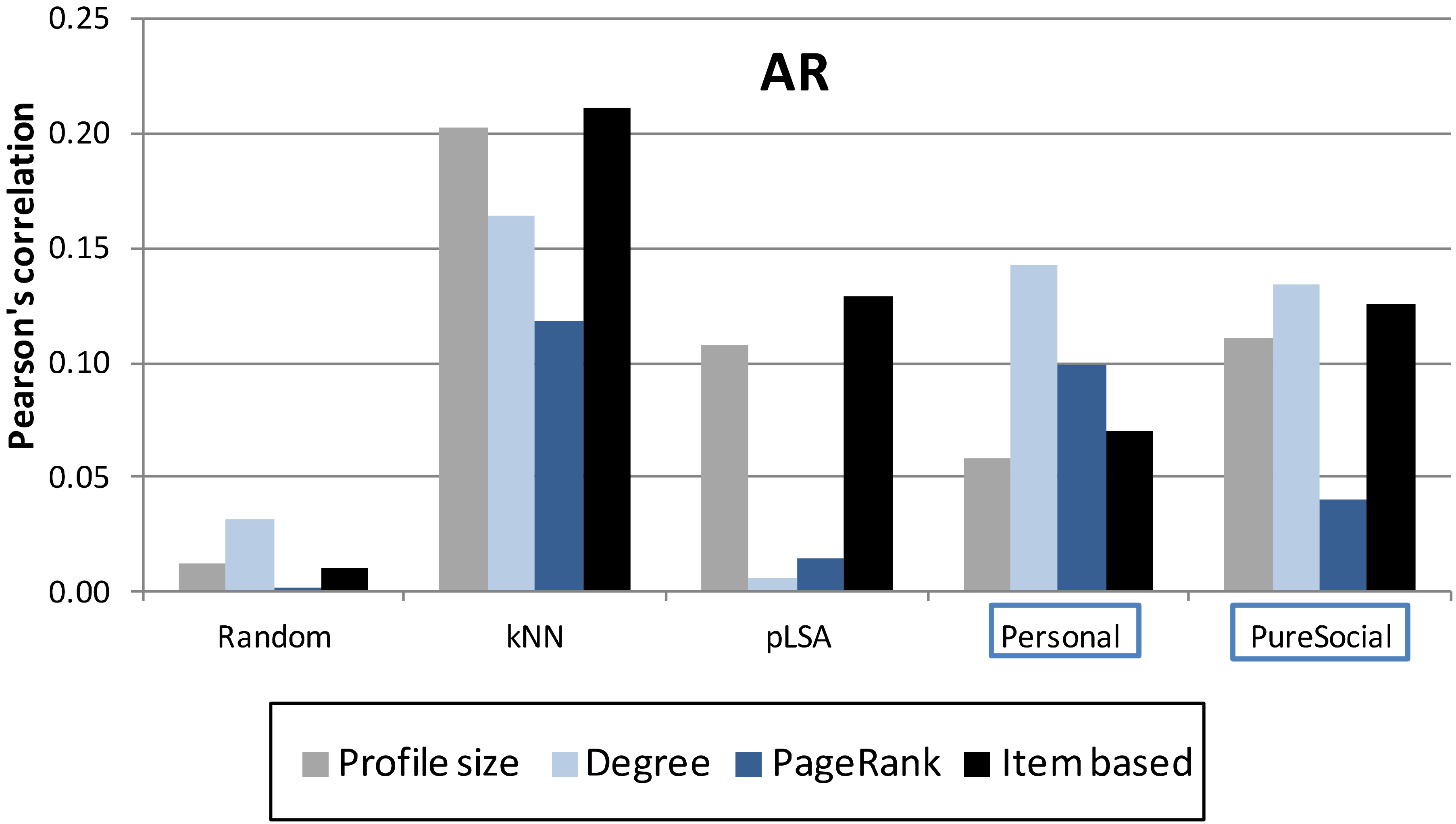
Experiments with log data

- **Temporal** and **frequency-based clarity predictors** show higher correlations than non-temporal predictors



Experiments with social data

- **Social predictors** have stronger correlations than rating predictors with social filtering recommenders (Personal and PureSocial)



Conclusions

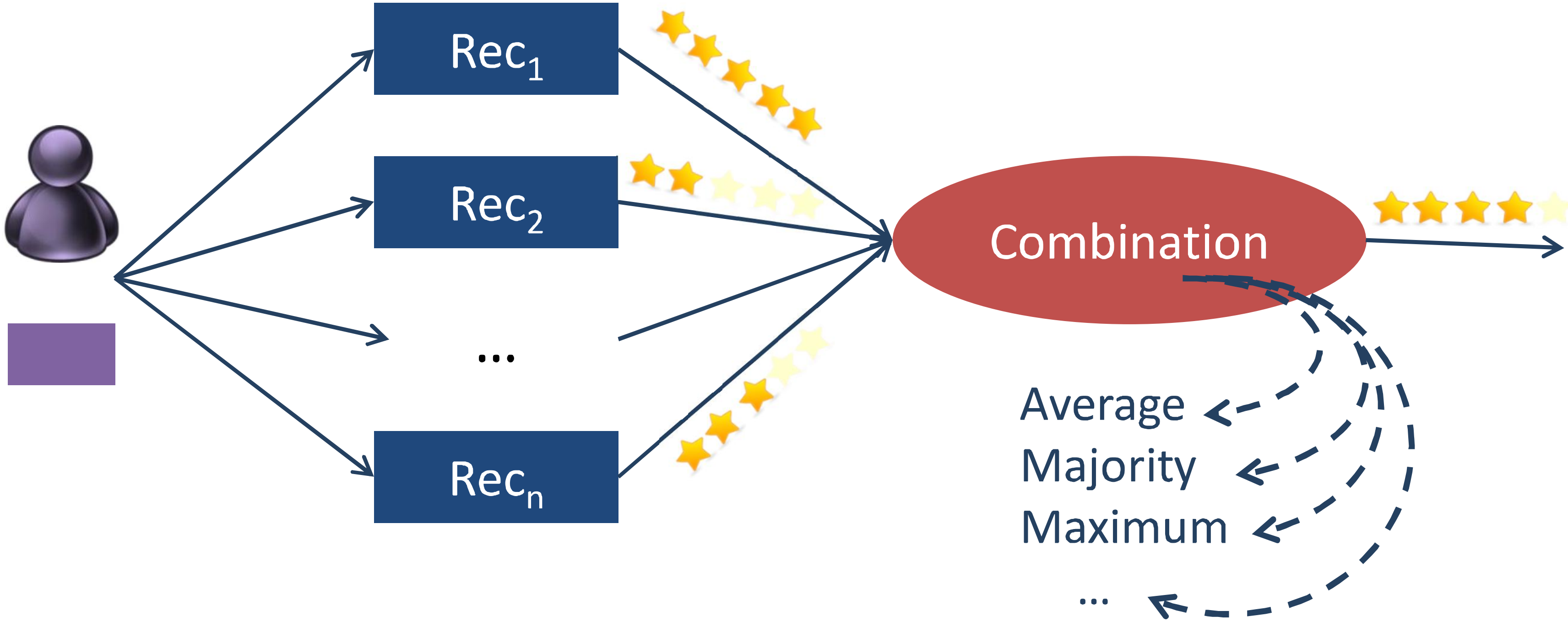
- **Strong predictive power** of the proposed predictors
  - Sanity check: **stronger correlations** than trivial predictors (e.g., profile size)
  - Better results than prediction based on training performance
- The **item based clarity** predictor consistently shows high correlation values in the three datasets evaluated
- **Correlations remain stable** with other evaluation metrics (nDCG and recall) and correlation coefficients (Spearman and Kendall)

Contents

Dynamic recommender ensembles

Dynamic recommender ensembles (1)

- Context
  - Hybrid recommendations are produced by combining the output of some recommenders
  - The combination of recommenders usually achieves better performance than separate methods
- Recommender ensembles



Dynamic recommender ensembles (1)

- Context
  - Hybrid recommendations are produced by combining the output of some recommenders
  - The combination of recommenders usually achieves better performance than separate methods
- Recommender ensembles (linear combination)

$$\tilde{r}(u,i) = \sum_k \lambda_k \cdot \tilde{r}_{R_k}(u,i) \quad \text{s.t.} \quad \sum_k \lambda_k = 1$$

- Research problem:

How to properly select the combination weights λ_k

Dynamic recommender ensembles (2)

- We propose to build dynamic ensembles (of size 2):

$$\tilde{r}(u,i) = \lambda_{R_1}(u,i) \cdot \tilde{r}_{R_1}(u,i) + \lambda_{R_2}(u,i) \cdot \tilde{r}_{R_2}(u,i)$$

  - The combination parameter depends on both the user and item
  - We use the performance predictors to assign these weights

- We assign the weight of R_1 according to the output of predictor $\gamma(u)$:
  - The weight of R_2 is fixed:

$$\tilde{r}(u,i) = \frac{\gamma(u)}{\gamma(u) + 0.5} \cdot \tilde{r}_{R_1}(u,i) + \frac{0.5}{\gamma(u) + 0.5} \cdot \tilde{r}_{R_2}(u,i)$$

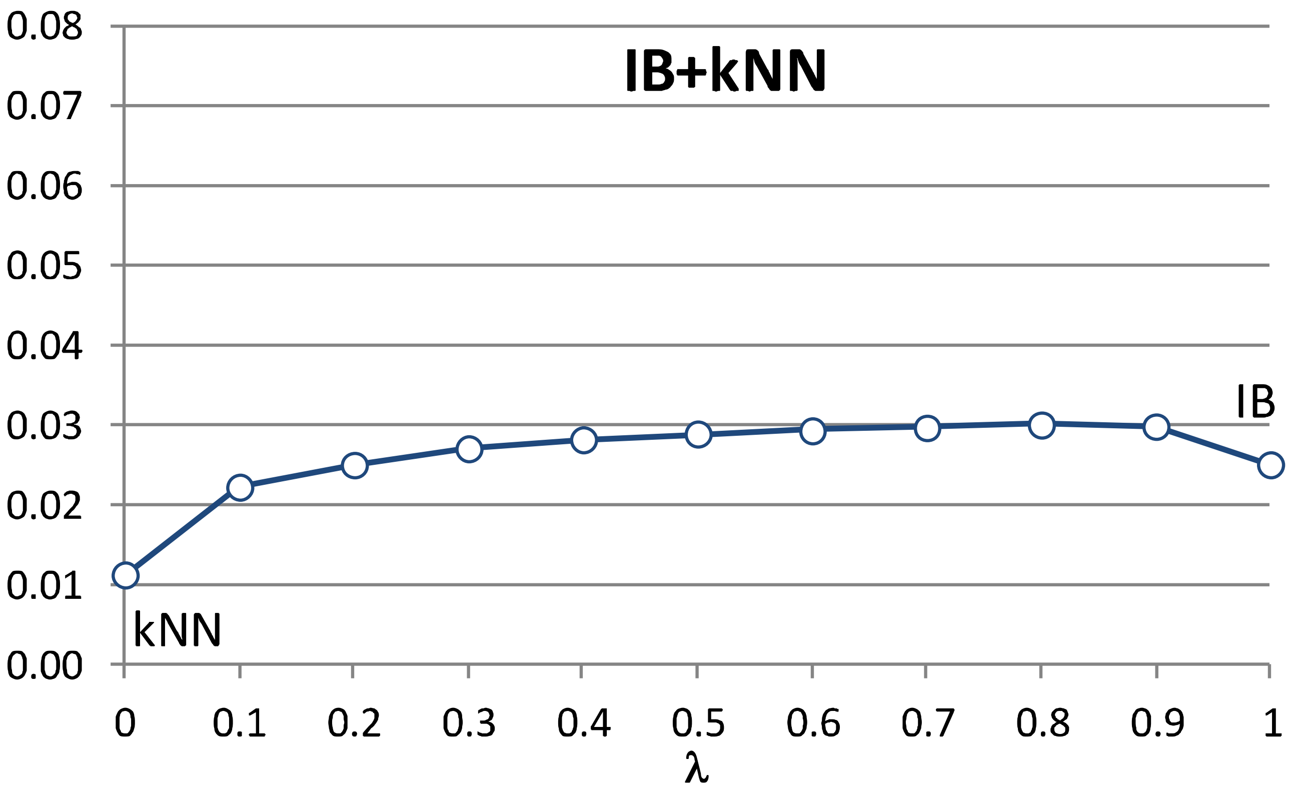
  - Or it depends on the predictor:

$$\tilde{r}(u,i) = \gamma(u) \cdot \tilde{r}_{R_1}(u,i) + (1 - \gamma(u)) \cdot \tilde{r}_{R_2}(u,i)$$

Requirements (1)

- Requirements for the problem to be well defined
  - Similar performance of the recommenders in the ensemble

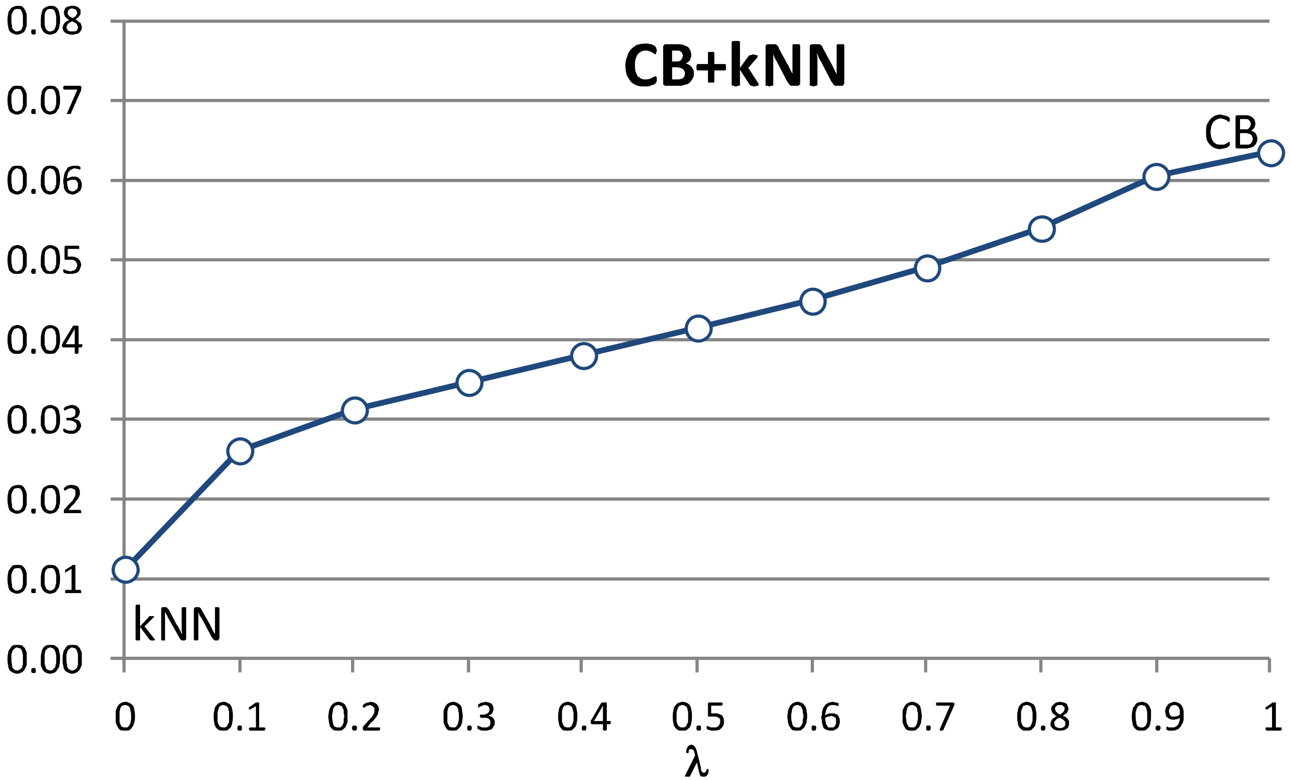
$$\tilde{r}(u,i) = \lambda \cdot \tilde{r}_{R_1}(u,i) + (1-\lambda) \cdot \tilde{r}_{R_2}(u,i)$$



Requirements (1)

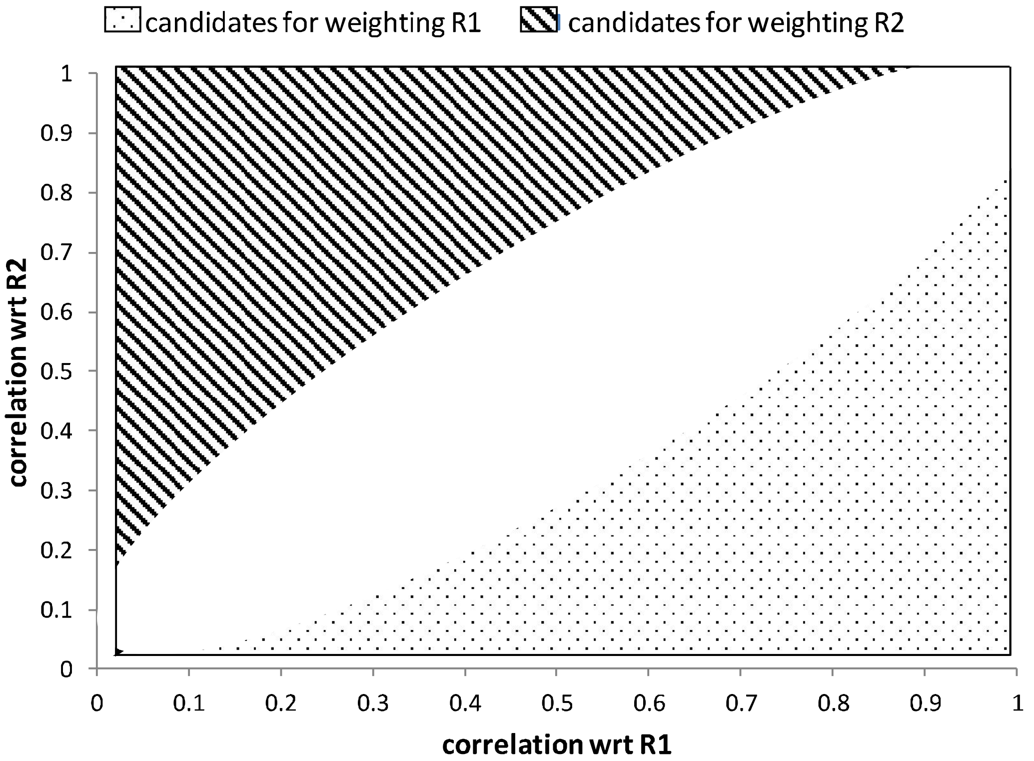
- Requirements for the problem to be well defined
  - Similar performance of the recommenders in the ensemble

$$\tilde{r}(u,i) = \lambda \cdot \tilde{r}_{R_1}(u,i) + (1-\lambda) \cdot \tilde{r}_{R_2}(u,i)$$



Requirements (2)

- Requirements for the problem to be well defined
  - Similar performance of the recommenders in the ensemble
- Requirements for our approach to be well defined
  - Positive correlation with one of the recommenders and neutral (or contrary) correlation with the other



Experiments

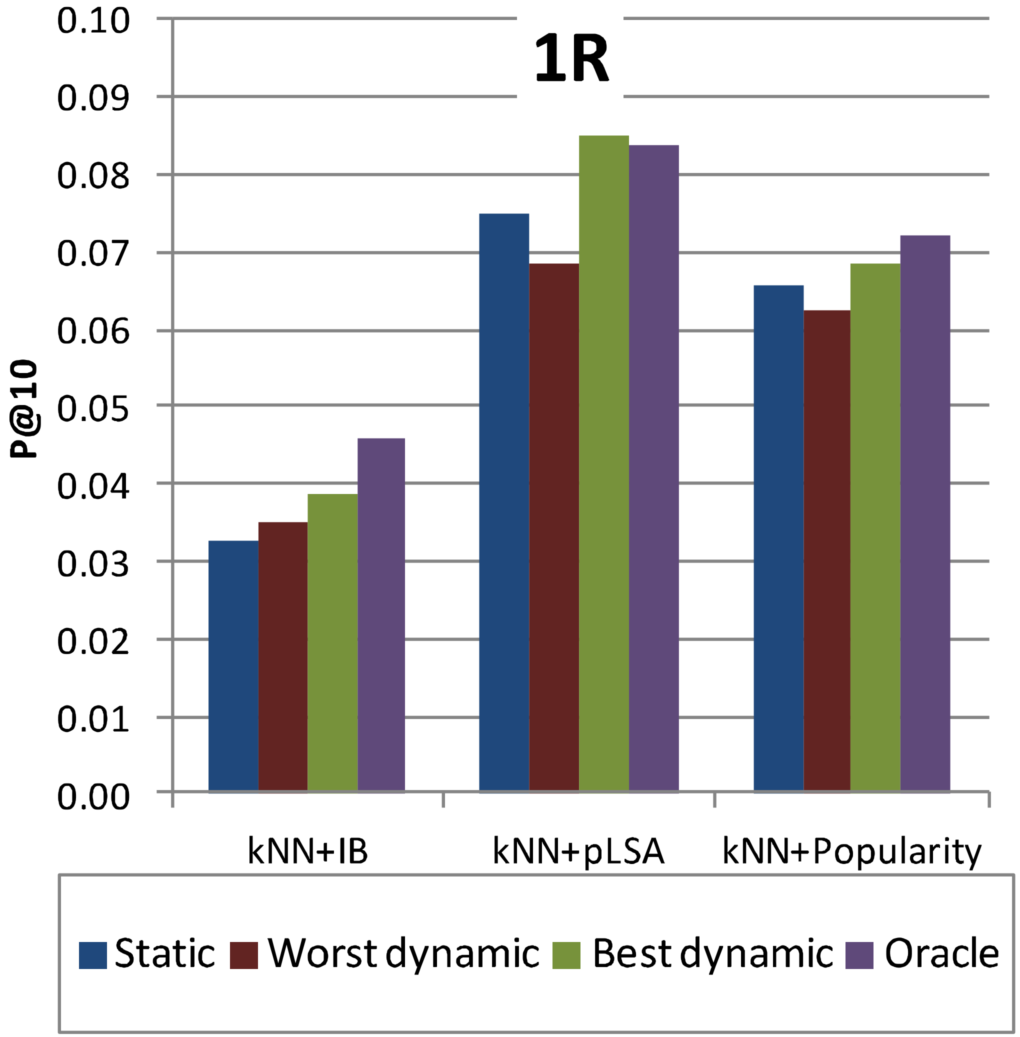
- Goal

Check if dynamic ensembles perform better than static ensembles

- Weighting schemes for $R_1 + R_2$
  - Static: same weight (0.5) for both recommenders and every user
  - Dynamic: weights from predictor's output (best and worst result)
  - Oracle: use weights from the true performance (perfect correlation)
- Metrics:
  - Precision at 10
- Evaluation methodologies
  - AR, 1R, P1R, U1R
- Datasets
  - MovieLens (ratings), Last.fm (logs), CAMRa (social)

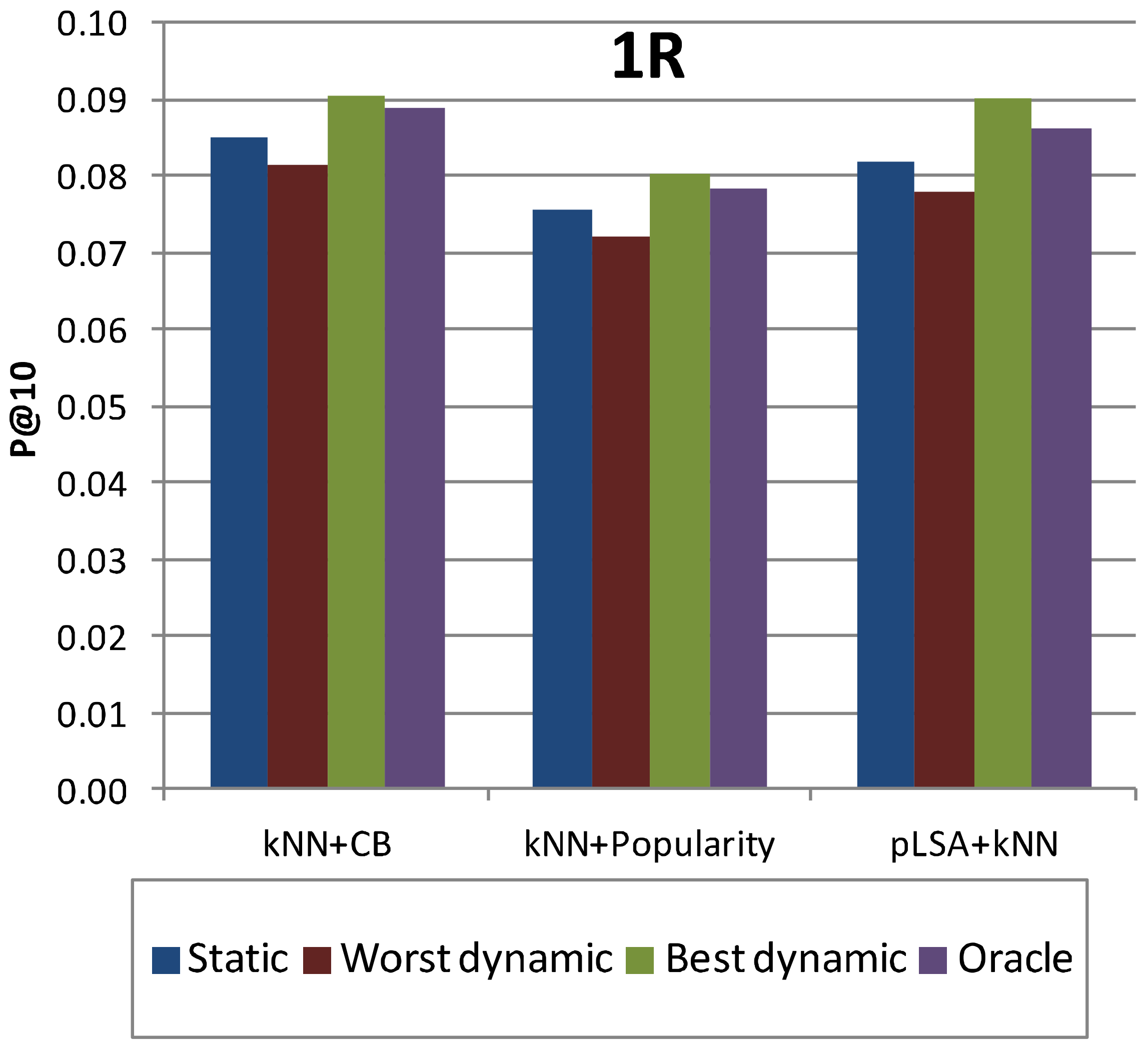
Experiments with rating data

- Dynamic ensembles perform better than the baseline
  - Similar results with AR and U1R, not so clear improvements with P1R



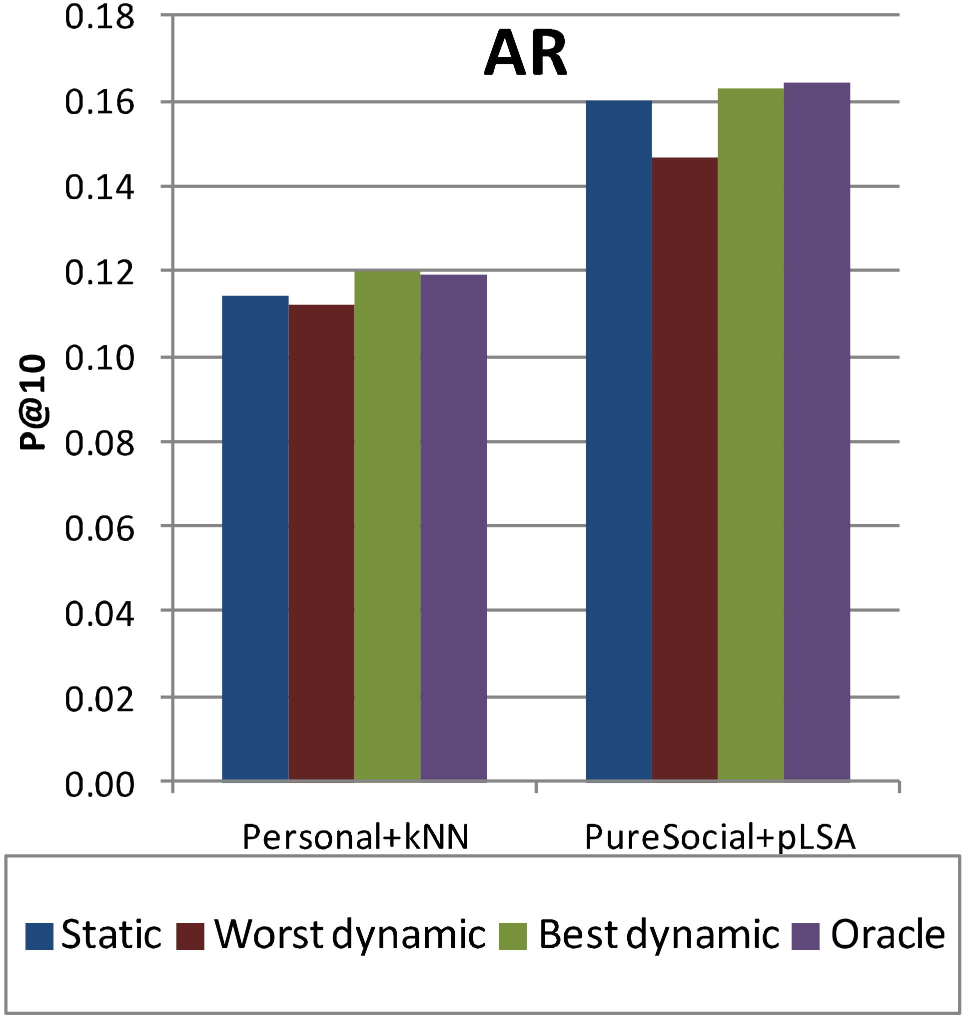
Experiments with log data

- Dynamic ensembles always outperform the baseline
- Better results than oracle



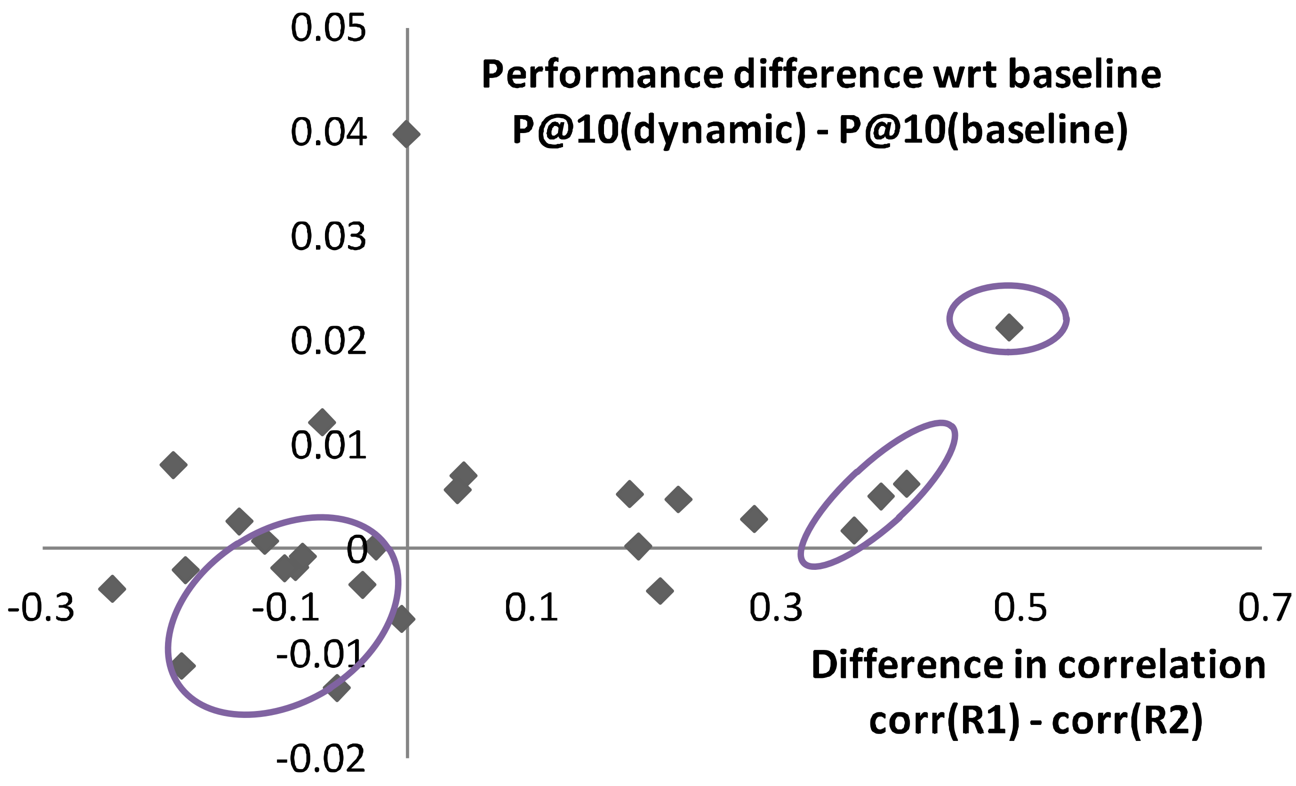
Experiments with social data

- Results less significative than before
- Due to lack of coverage, 1R does not provide sensible results



Summary of results

- The larger the difference in correlation, the better the improvement over the baseline
  - The following is validated: "correlations with each recommender should not be very similar"



Neighbour selection and weighting in Collaborative Filtering

Neighbour selection and weighting in CF

- User-based collaborative filtering:

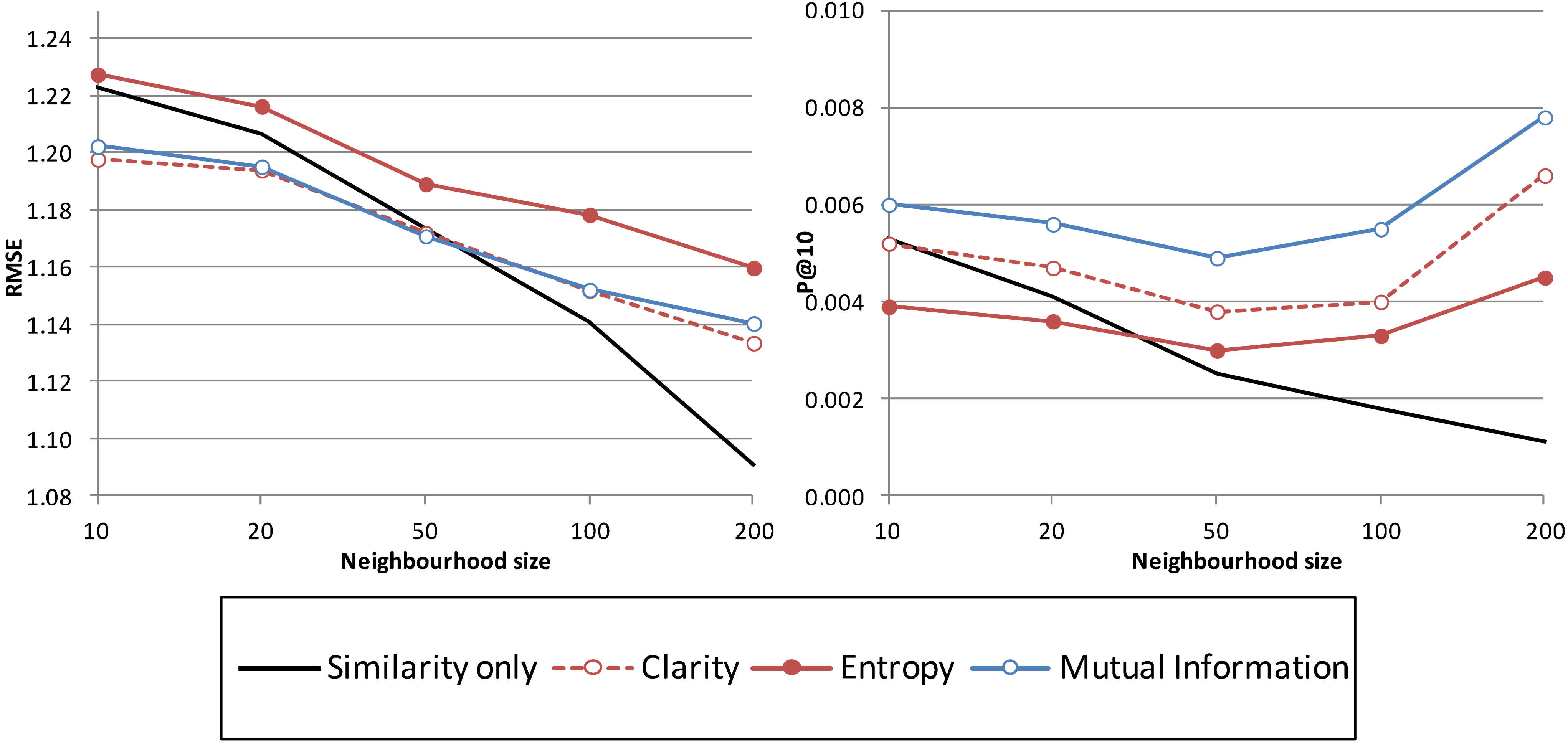
$$\tilde{r}(u,i) = \bar{r}(u) + C \sum_{v \in V} \mathrm{sim}(u,v)\big(r(v,i) - \bar{r}(v)\big)$$

- Use <u>neighbour performance predictors</u> (function γ) to **select** and **weight** neighbours' contribution to the recommendations

$$\tilde{r}(u,i) = \bar{r}(u) + C \sum_{v \in f^{neigh}(u,i;k,\gamma)} f^{agg}\big(\gamma(u,v,i), \mathrm{sim}(u,v)\big)\big(r(v,i) - \bar{r}(v)\big)$$

Results

- Performance improvement in both RMSE and Precision
  - For RMSE: better (lower values) for smaller neighbourhoods
  - For Precision: better (higher values) with larger neighbourhoods



Contents

RG1: Evaluating performance in recommender systems

Conclusions – RG1

- **Assumptions and conditions** underlying IR evaluation methodologies **are not granted** in usual recommendation settings
- We detect **statistical biases** in evaluation of recommender systems: sparsity and popularity
- We **propose novel experimental approaches** that neutralise the popularity bias

RG2: Predicting performance in recommender systems

Conclusions – RG2

- We define **performance predictors** for recommendation, with several variations of user clarity
- We integrate the **temporal** and **social dimensions**
- We find predictors with **significant predictive power**, also under unbiased conditions, that is, when sparsity and popularity biases have been neutralised

RG3: Applications

Conclusions – RG3

- We **aggregate** the output of recommenders and neighbours **using performance predictors**
- We define a dynamic hybrid framework where **high correlation values with performance tend to correspond with enhancements in dynamic ensembles**
- We propose a framework for neighbour selection and weighting unifying several notions of neighbour performance where we obtained **improvements in terms of RMSE and precision**

Future work

- RG1: *Evaluating performance in recommender systems*
  - Extend our analysis on design alternatives to other ranking metrics (e.g., AUC)
  - Validate the unbiased methodologies with online evaluations
- RG2: *Predicting performance in recommender systems*
  - Combine predictors to obtain higher correlation values
  - Use clustering approaches to estimate the quality of predictors
- RG3: *Applications*
  - Extend the experiments with ensembles of N recommenders and using one predictor for each recommender
  - Adapt the proposed neighbour performance metrics to use ranking metrics

Thank you!

Performance prediction and evaluation in Recommender Systems: An Information Retrieval Perspective

Alejandro Bellogín Kouki

under the supervision of

Pablo Castells Azpilicueta

and

Iván Cantador Gutiérrez

Publications (1)

- Journals

1. Bellogín, A., Wang, J., and Castells, P. Bridging Memory-Based Collaborative Filtering and Text Retrieval. *Information Retrieval Journal*, to appear.
2. Bellogín, A., Cantador, I., and Castells, P. A Comparative Study of Heterogeneous Item Recommendations in Social Systems. *Information Sciences*, to appear.
3. Bellogín, A., Cantador, I., Díez, F., Castells, P., and Chavarriaga, E. (2012). An empirical comparison of social, collaborative filtering, and hybrid recommenders. *ACM Transactions on Intelligent Systems and Technology*, to appear.
4. Cantador, I., Castells, P., and Bellogín, A. (2011). An enhanced semantic layer for hybrid recommender systems. *International Journal on Semantic Web and Information Systems*, 7(1):44–78.
5. Cantador, I., Bellogín, A., and Castells, P. (2008). A multilayer ontology-based hybrid recommendation model. *AI Commun.*, 21(2-3):203–210.

Publications (2)

- Conferences

1. Campos, P. G., Bellogín, A., Díez, F., and Cantador, I. (2012). Time Feature Selection for Identifying Active Household Members. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, CIKM '12, New York, NY, USA. ACM (to appear).
2. Bellogín, A. and Parapar, J. (2012). Using Graph Partitioning Techniques for Neighbour Selection in User-Based Collaborative Filtering. In *Proceedings of the sixth ACM conference on Recommender systems*, RecSys '12, pages 213–216, New York, NY, USA. ACM. Best short paper award
3. Bellogín, A., Wang, J., and Castells, P. (2011). Structured collaborative filtering. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, pages 2257–2260, New York, NY, USA. ACM.
4. Bellogín, A., Castells, P., and Cantador, I. (2011). Precision-oriented evaluation of recommender systems: an algorithmic comparison. In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 333–336, New York, NY, USA. ACM.
5. Bellogín, A., Castells, P., and Cantador, I. (2011). Predicting the Performance of Recommender Systems: An Information Theoretic Approach. In Amati, G. and Crestani, F., editors, *ICTIR*, volume 6931 of *Lecture Notes in Computer Science*, pages 27–39, Berlin, Heidelberg. Springer Berlin / Heidelberg.
6. Bellogín, A., Castells, P., and Cantador, I. (2011). Self-adjusting hybrid recommenders based on social network analysis. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, SIGIR '11, pages 1147–1148, New York, NY, USA. ACM.
7. Bellogín, A., Wang, J., and Castells, P. (2011). Text Retrieval Methods for Item Ranking in Collaborative Filtering. In Clough, P., Foley, C., Gurrin, C., Jones, G., Kraaij, W., Lee, H., and Mudoch, V., editors, *Advances in Information Retrieval*, volume 6611 of *Lecture Notes in Computer Science*, chapter 30, pages 301–306. Springer Berlin / Heidelberg, Berlin, Heidelberg.
8. Cantador, I., Bellogín, A., and Vallet, D. (2010). Content-based recommendation in social tagging systems. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 237–240, New York, NY, USA. ACM.
9. Bellogín, A. and Castells, P. (2010). A Performance Prediction Approach to Enhance Collaborative Filtering Performance. In Gurrin, C., He, Y., Kazai, G., Kruschwitz, U., Little, S., Roelleke, T., Rüger, S., and Rijsbergen, editors, *Advances in Information Retrieval*, volume 5993 of *Lecture Notes in Computer Science*, pages 382–393–393, Berlin, Heidelberg. Springer Berlin / Heidelberg.
10. Bellogín, A. and Castells, P. (2009). Predicting Neighbor Goodness in Collaborative Filtering. In *8th International Conference on Flexible Query Answering Systems (FQAS 2009)*. Roskilde, Denmark, pages 605–616. Springer Verlag Lecture Notes in Computer Science.
11. Cantador, I., Bellogín, A., and Castells, P. (2008). News@hand: A Semantic Web Approach to Recommending News. In Nejdl, W., Kay, J., Pu, P., and Herder, E., editors, *Adaptive Hypermedia and Adaptive Web-Based Systems*, volume 5149 of *Lecture Notes in Computer Science*, chapter 34, pages 279–283. Springer Berlin / Heidelberg, Berlin, Heidelberg.