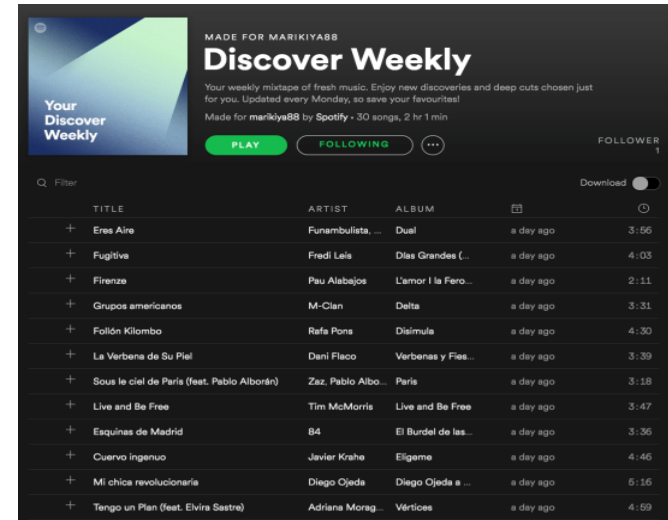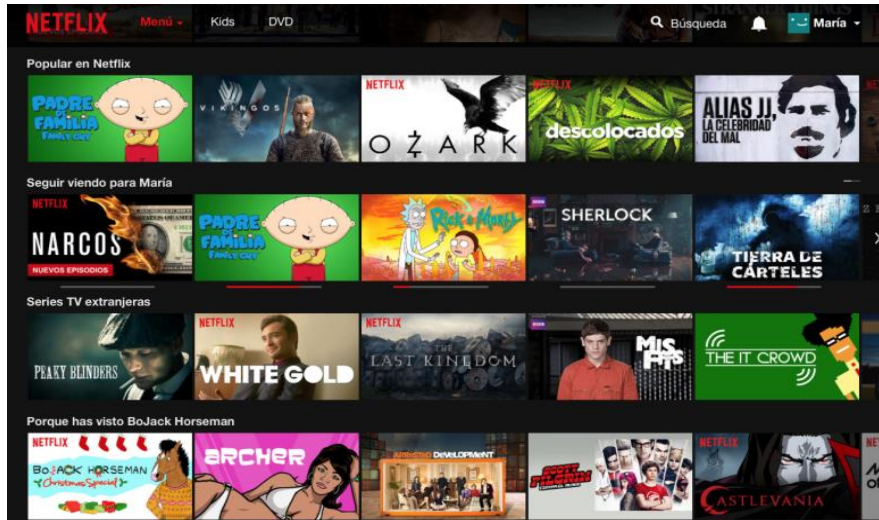# Aspect-based active learning for user preference elicitation in recommender systems

María Hernández Rubio (*presenting author*)
Alejandro Bellogín
Iván Cantador

Escuela Politécnica Superior

UNIVERSIDAD AUTONOMA DE MADRID

Recommender systems

User Preferences

Ratings

Intex 86" x 23" Rectangular Frame Above Ground Outdoor Child Safe Splash Swimming Pool
by Intex
★★★★☆ ⌄   4,473 ratings
Available from these sellers.
Size: 7' 2" x 4' 11" x 23"

| 7' 2" x 4' 11" x 23" | 8' 6" x 5' 3" x 25" | 14' 9" x 7' 3" x 33" |
| 9 options from $326.50 | 1 option from $488.88 | 3 options from $699.00 |

- SUMMERTIME FUN: Escape the summer heat and help keep your little ones cool with the Intex Rectangular Frame Baby Pool. This pool can fit up to 6 people and has a water capacity of 439 gallons of water.

Categorical

Thumbs up / down

👍 6 K   👎 186

Reviews

jwood1964

★★★★★ PERFECT FOR BACK PATIO
Reviewed in the United States on May 3, 2019
Size: 8' 6" x 5' 3" x 25"   Verified Purchase

Easy to set up, Instructions were clear. I am seriously thinking about buying another one for Camping! Enough space for water exercise. We added a pump/filter and it creates a perfect back yard oasis. We used foam (for under laminate flooring) as underlay. very comfortable. An adult can sit in the pool and water is up to shoulders. There is a cover sold on amazon which is a perfect fit! Bestway 58105 Frame Pool Cover, 102 by 67-Inch and the Intex filter/pump is Intex Krystal Clear Cartridge Filter Pump for Above Ground Pools, 1000 GPH Pump Flow Rate, 110-120V with GFCI also sold on amazon!

Aspect Information

User preferences acquisition

Preference elicitation: how to model user's preferences

➡ **Active Learning (AL):** ask users to rate items smartly
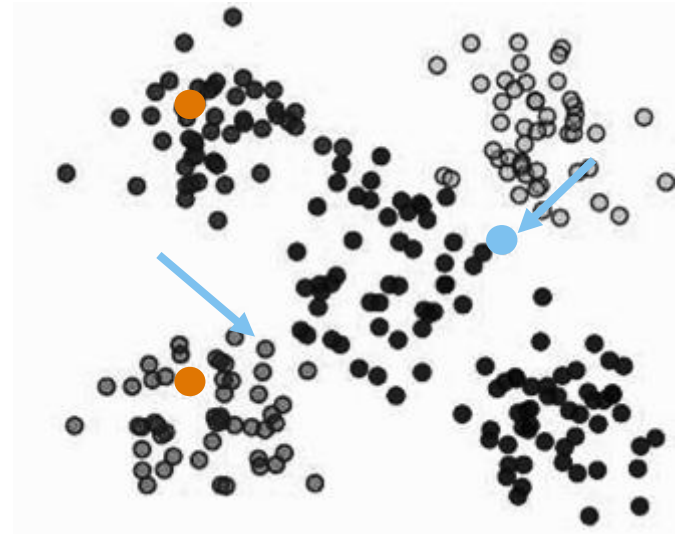
Our work

Build an **AL algorithm** based on **aspect opinions** extracted from **reviews**.

Objective: get similar recommendation metrics with fewer item asked.

Table of contents

Escuela Politécnica Superior

UNIVERSIDAD AUTONOMA DE MADRID

Table of contents

Escuela Politécnica Superior

UAM
UNIVERSIDAD AUTONOMA
DE MADRID

Mehdi et al. TIST (2013)

SoA item-based methods

- **Non-Personalized** vs **Personalized** Active Learning:
  - Take into account users' previously expressed ratings
  - Request all the users to rate the same items

- **Single-** vs **combined-heuristics**
  - Single: implements a unique item selection rule
  - Combined: hybridize several single-heuristics strategies

Mehdi et al. TIST (2013)
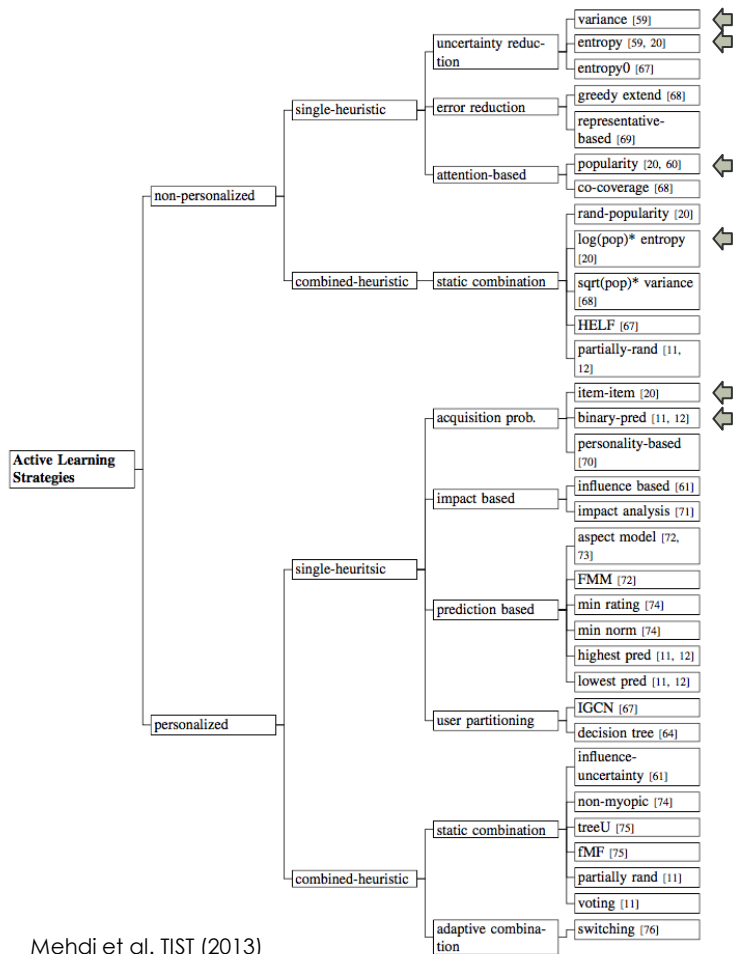
- **variance**: items with highest rating variance
- **popularity**: items with highest number of ratings
- **entropy**: items with highest rating dispersion
- **log(pop)*entropy**

- **item-item**: items more similar to user's previously rated items
- **binary-pred**: items with highest probability of being rated by the user

Table of contents

Escuela Politécnica Superior

UNIVERSIDAD AUTONOMA DE MADRID

Aspect-based Active Learning method

*Exploiting the rich information that can be extracted from reviews: **item aspects** mentioned and the **opinion or sentiment** associated to them.*

- Help user to find items that share characteristics with previously interacted items

- Item aspects (vs other content or collaborative information) should alleviate the cold-start problem

Escuela Politécnica Superior

UA
UNIVERSIDAD AUTONOMA
DE MADRID

Aspect-based Active Learning method

*Exploiting the rich information that can be extracted from reviews: **item aspects** mentioned and the **opinion or sentiment** associated to them.*

- Hybrid recommendation approach (Frolov & Oseledets, RecSys 2019): aspect-based item-item similarity matrix plus collaborative information.

- Similarity between item i_n and i_m is computed as the cosine similarity over the item profile $i_n = \{w_{na}\}^K_{a=1}$ built on the K aspect opinions, where w_{na} is the weight assigned to aspect a for item i_n.

$$\mathrm{sim}(i_n, i_m) = \frac{\sum w_{na} w_{ma}}{\sqrt{\sum w_{na}^2 w_{ma}^2}}$$

- *item-item* (personalized and single heuristic)

Escuela Politécnica Superior

Table of contents

Escuela Politécnica Superior

UNIVERSIDAD AUTONOMA DE MADRID

Dataset

- Product Dataset: Movies & TV Amazon product reviews dataset (McAuley, WWW (2016))

- Aspect method: vocabulary (voc) (Hernández-Rubio et al. UMUAI (2019))

	Ratings	Users	Items	Annotations	Aspects
Initial	1,697,533	123,960	50,052	369,175	23
Items with aspects	1,683,190	123,960	48,074	369,175	23
Users with >= 20 ratings	819,148	14,010	47,506	367,750	23

Table 1: dataset and aspects statistics

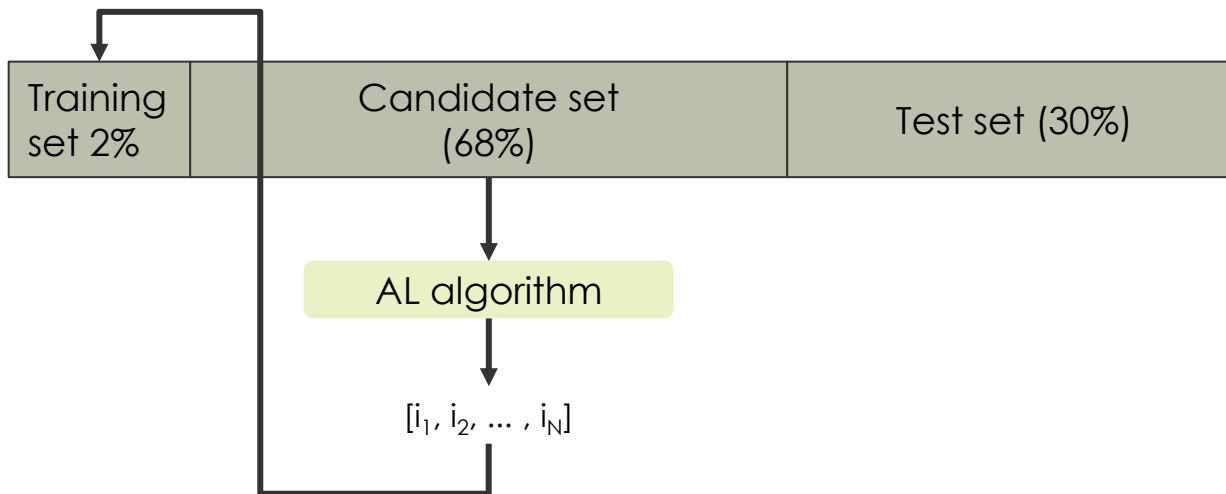** for this work we have sample to 1500 users for computational reasons*

Evaluation

Methodology:

Training set 2%	Candidate set (68%)	Test set (30%)

Escuela Politécnica Superior

UNIVERSIDAD AUTONOMA DE MADRID

Evaluation

Methodology:

Evaluation

Methodology:

Training set 2%		Candidate set (68%)	Test set (30%)

AL algorithm

$[i_1, i_2, ... , i_N]$

SVD

metrics

Escuela Politécnica Superior

UAM
UNIVERSIDAD AUTONOMA
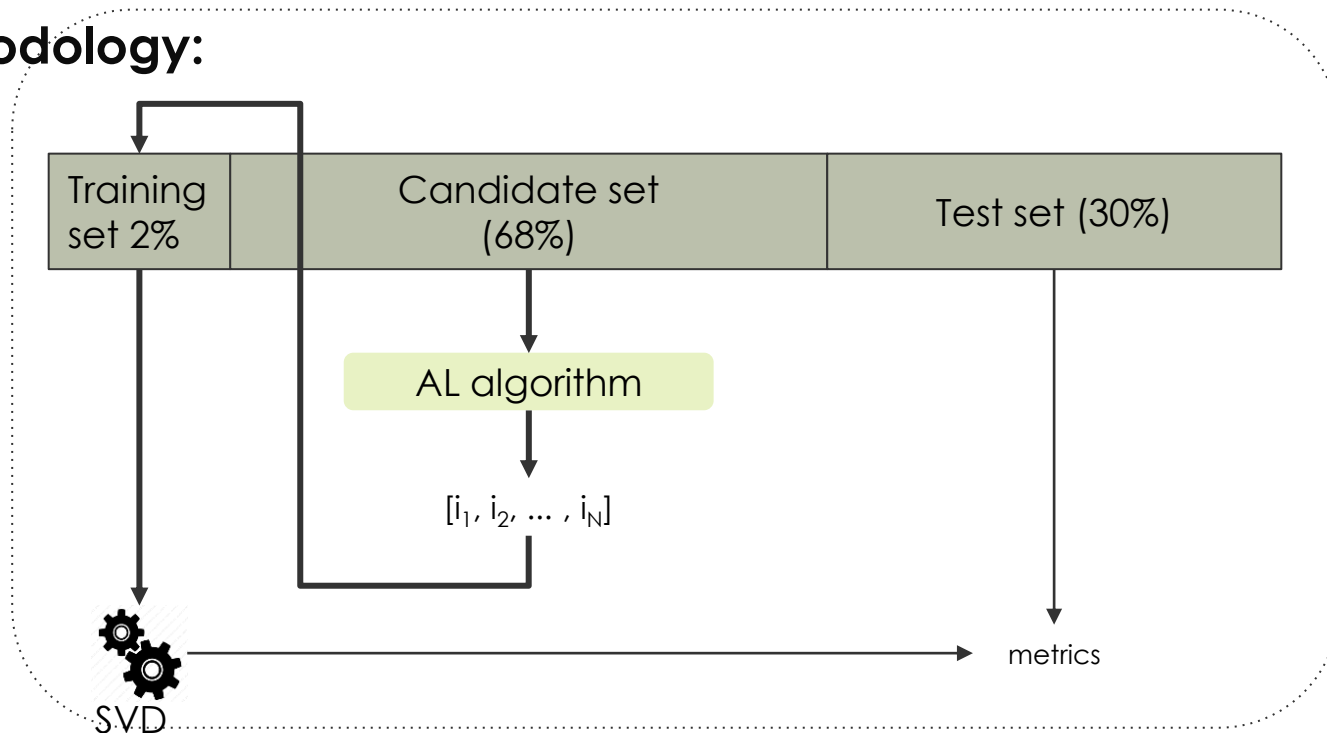DE MADRID
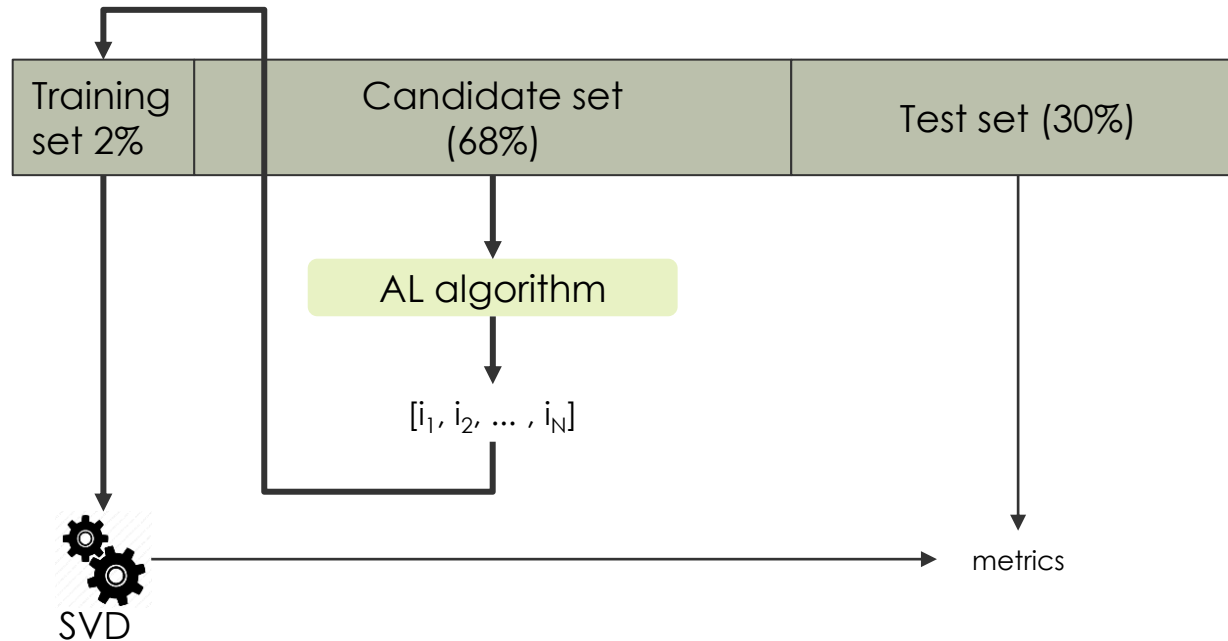
Evaluation

Methodology:

N = 10
iter = 170
CV = 3

Evaluation

- Metrics:
    - Rating: MAE, RMSE
    - Ranking: P@1, P@5, P@10

- Baselines:
    - random
    - variance
    - popularity
    - entropy
    - log-pop-entropy

Results

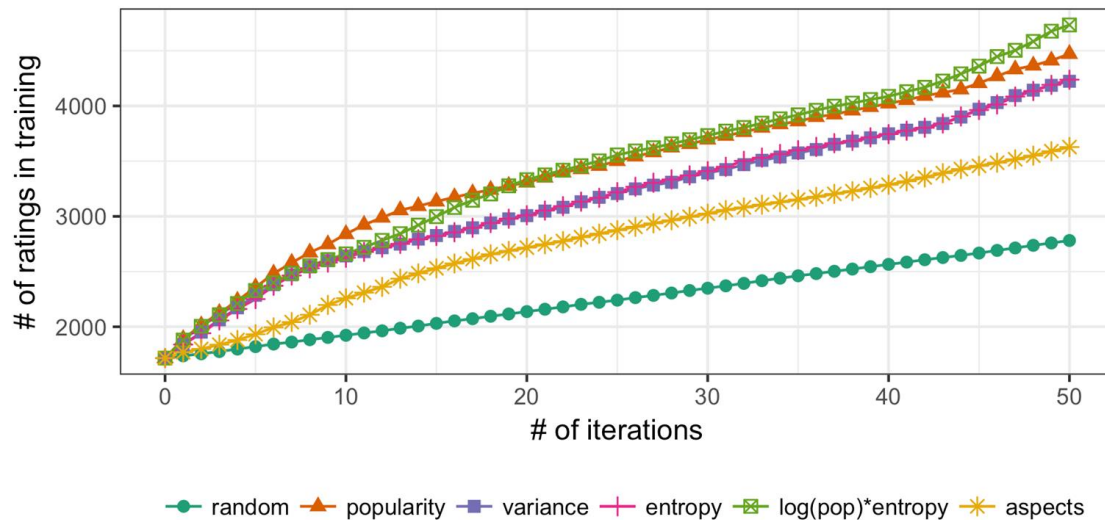- Aspect-based method is not able to find all known items for the user

Figure 1: Evolution on the number of ratings correctly elicited by each strategy (zoomed in on the first 50 iterations)

Results

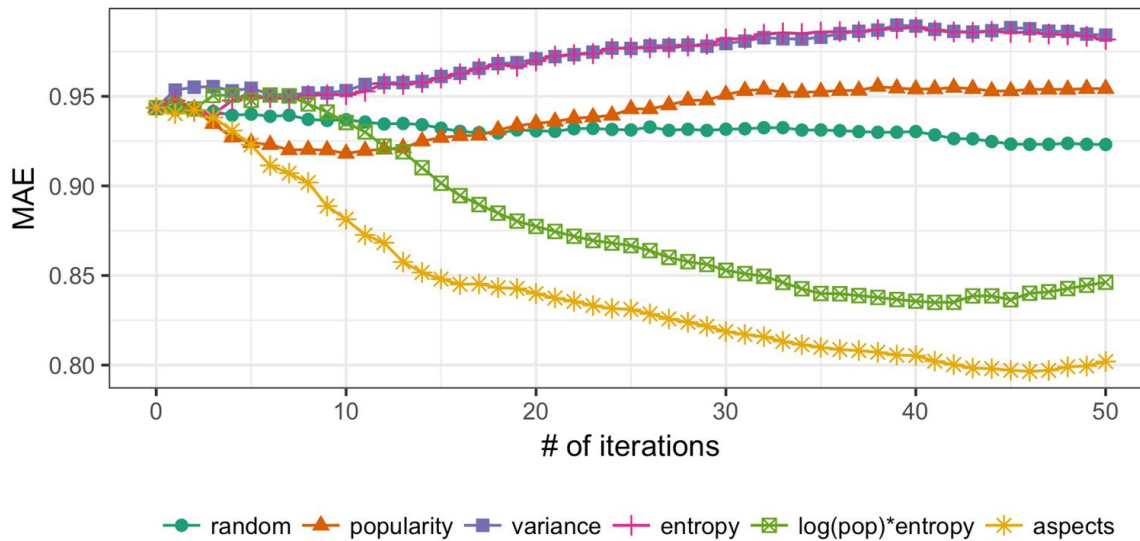- Aspect-based method gets the highest improvement in error.

Figure 2: Evolution on the error accuracy (the lower, the better) under the effect of six elicitation strategies.

Results

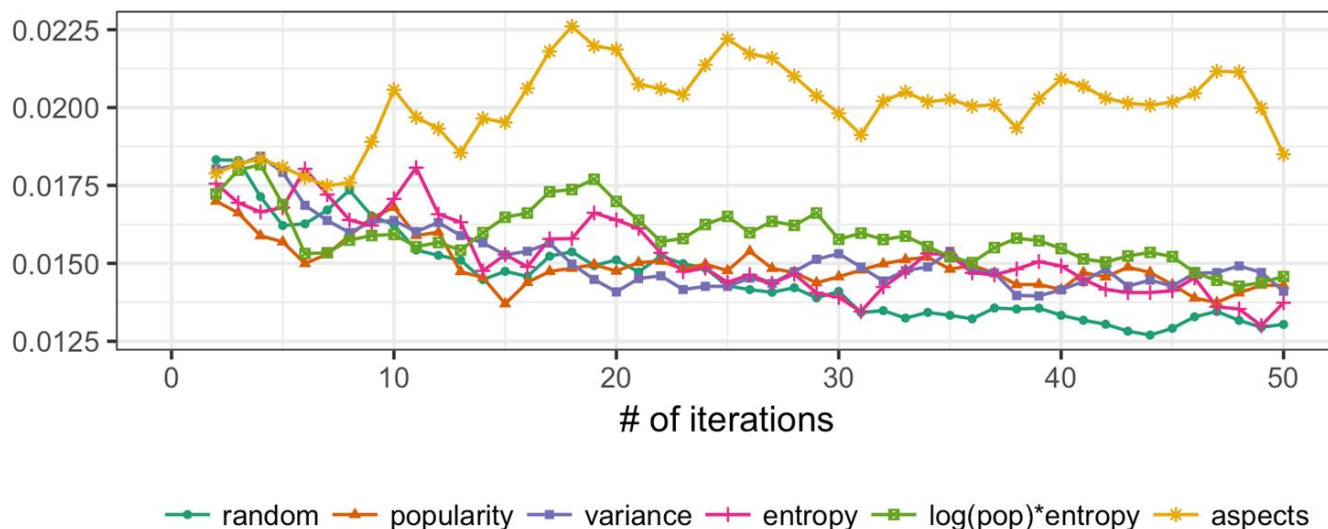- Aspect-based method is the best performing method throughout most of the elicitation process.

Figure 3: Ranking accuracy measured as P@5 (the higher, the better) under the effect of six elicitation strategies, smoothed values taking the average of the last 3 points.

Table of contents

Conclusions

- Novel active learning approach based on opinions about item aspects.

- Tested on a real world dataset

- Outperforms AL strategies on rating prediction error and ranking precision metrics.

Escuela Politécnica Superior

UNIVERSIDAD AUTONOMA
DE MADRID

Future Work

- More exhaustive experiments:
  - more sophisticated aspect extraction methods
  - several recommender systems
  - datasets from several domains

- Analyze the behaviour of our method on different cold-start settings

- Online evaluation with real users to confirm offline results

- Integrate into a conversational agent or chatbot

Questions?
Thank you!