



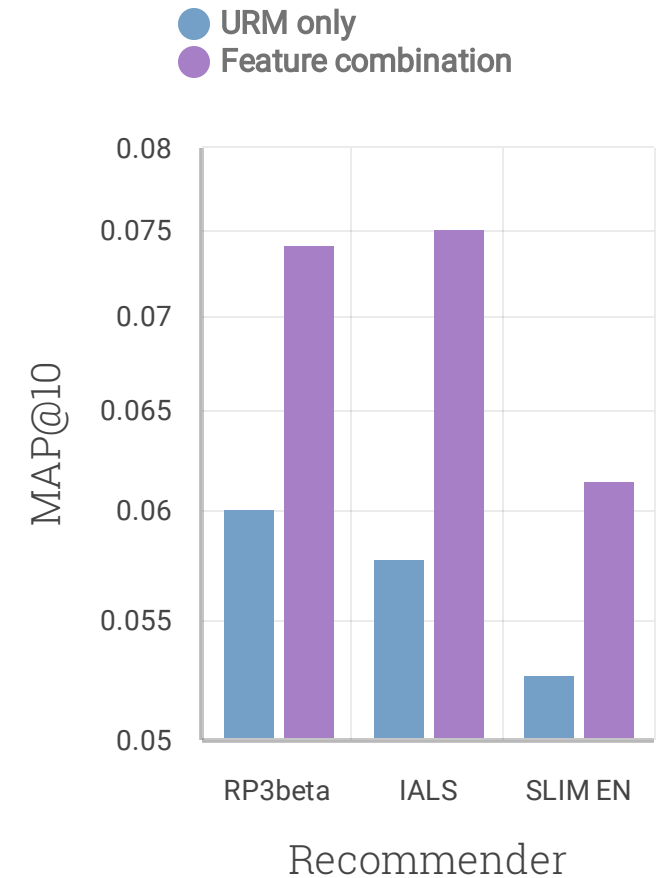
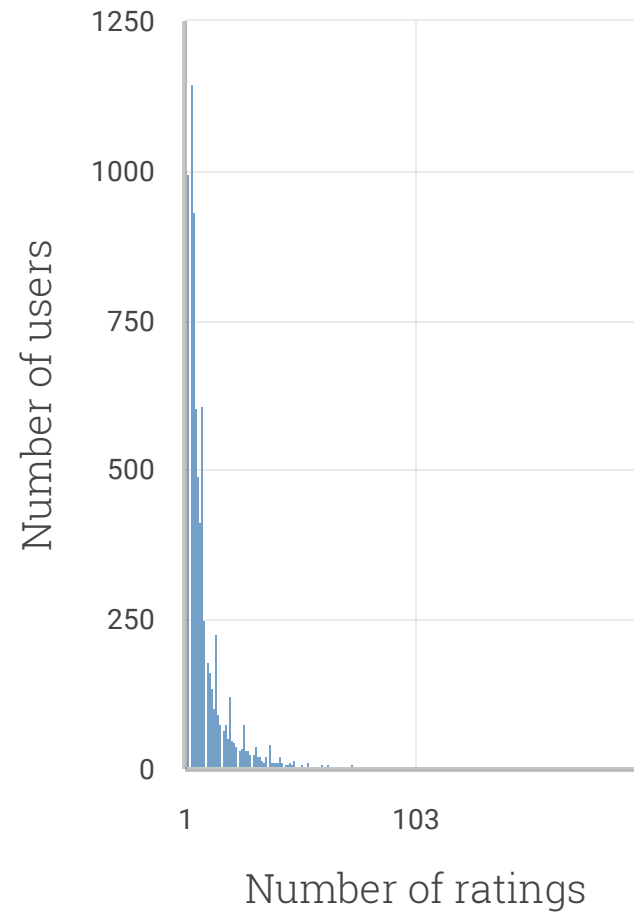
POLITECNICO
MILANO 1863

RecSys Challenge 2020

Alessandro Sanvito

Some considerations on data

- Most of the users have very few interactions
- The quality of recommendations declines for users with more than 6 interactions (the median)
- Feature combination considerably improves performances on the whole dataset



Hybrid weights and performances

Implementation details

- Hyperparameters tuned once on the whole dataset, then on 25% users with the most ratings (the long tail)
- All recommenders trained on URM+ICM
- Different confidences for the ICM and the URM in training IALS

Recommender	Weight	MAP@10	Weight (long tail)	MAP@10 (long tail)	MAP@10 (Kaggle average)
RP3beta	39%	0.0741	28%	0.0707	0.0939
IALS	46%	0.0751	41%	0.0728	N.A.
SLIM Elastic Net	15%	0.0615	31%	0.0596	N.A.
Merged Hybrid	100%	0.0804	100%	0.0777	0.1028



Some takeaways



What did not work

- Binarized ICM
- Pipelined hybrids
- Normalized ratings



What did work

- Keeping the model simple
- Bayesian optimization
- 5-fold cross validation
- MaurizioFD's repo
- ...and a LOT of trial and error



Further improvements

- RP5beta applied to the long tail
- Tuning the user split percentage
- Feature engineering





It's-a me, Alessandro

- **Data Science** MSc. student @EIT Digital Master School (POLIMI, KTH)
- **Computer Engineering** BSc. @Polimi
- @alexdruso on **Github** (and everywhere else)
- **Eager** to take part in the international challenge :)