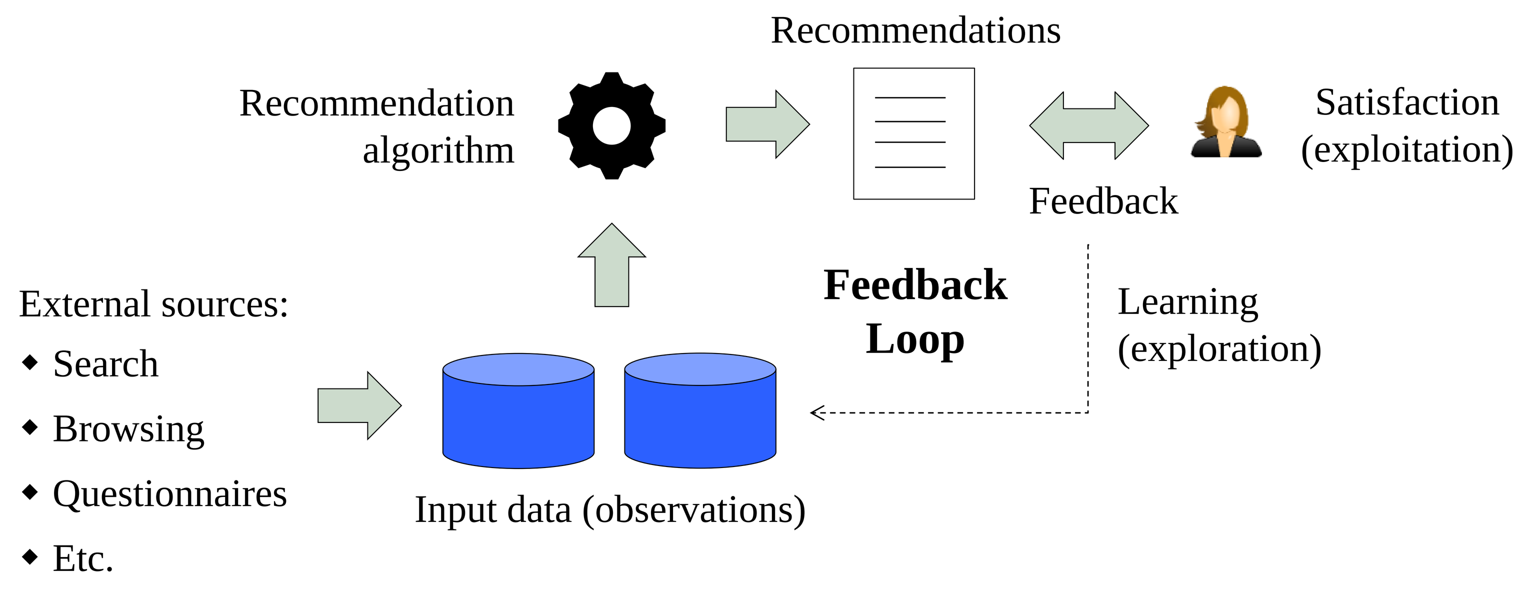
Multi-Armed Recommender System Bandit Ensembles

Rocío Cañamares, Marcos Redondo and Pablo Castells
Universidad Autónoma de Madrid
{rocio.cannamares,pablo.castells}@uam.es
marcos.redondo@estudiante.uam.es

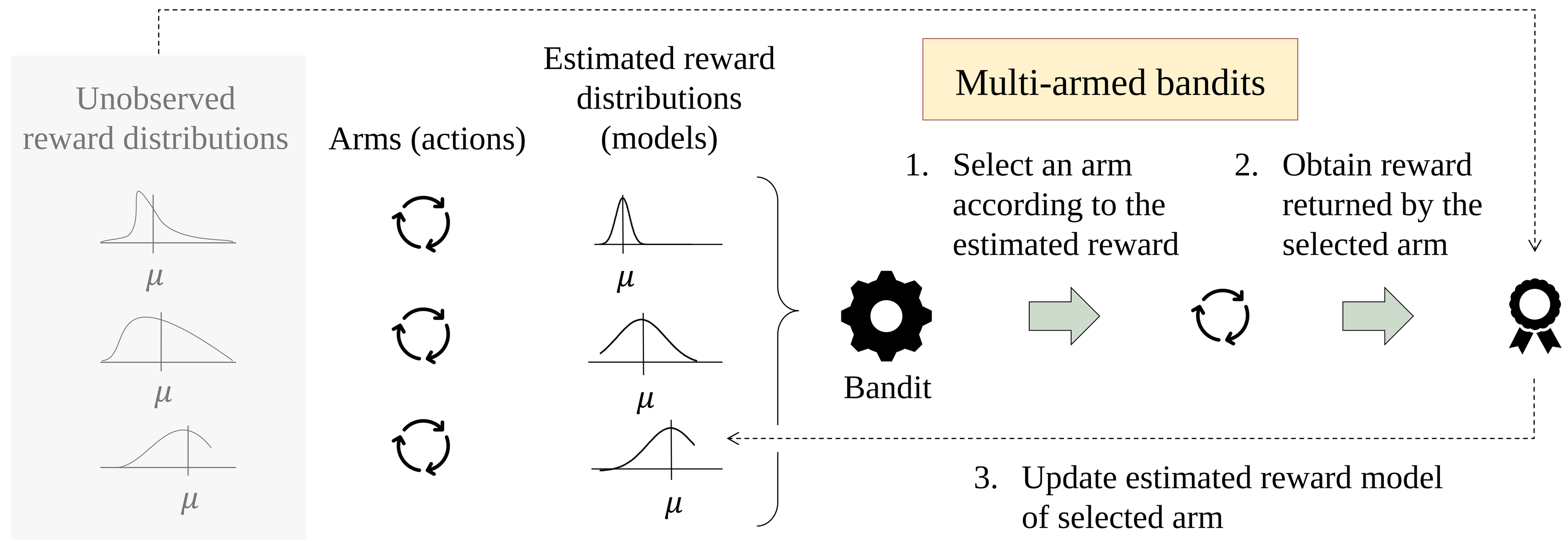
Universidad Autónoma de Madrid
IR Group @UAM

Bandit Recommender System Ensembles

Recommendation is a cyclic interactive process



Multi-armed bandits



Bandit Recommender System Ensembles

Arms ⟷ Recommendation algorithms

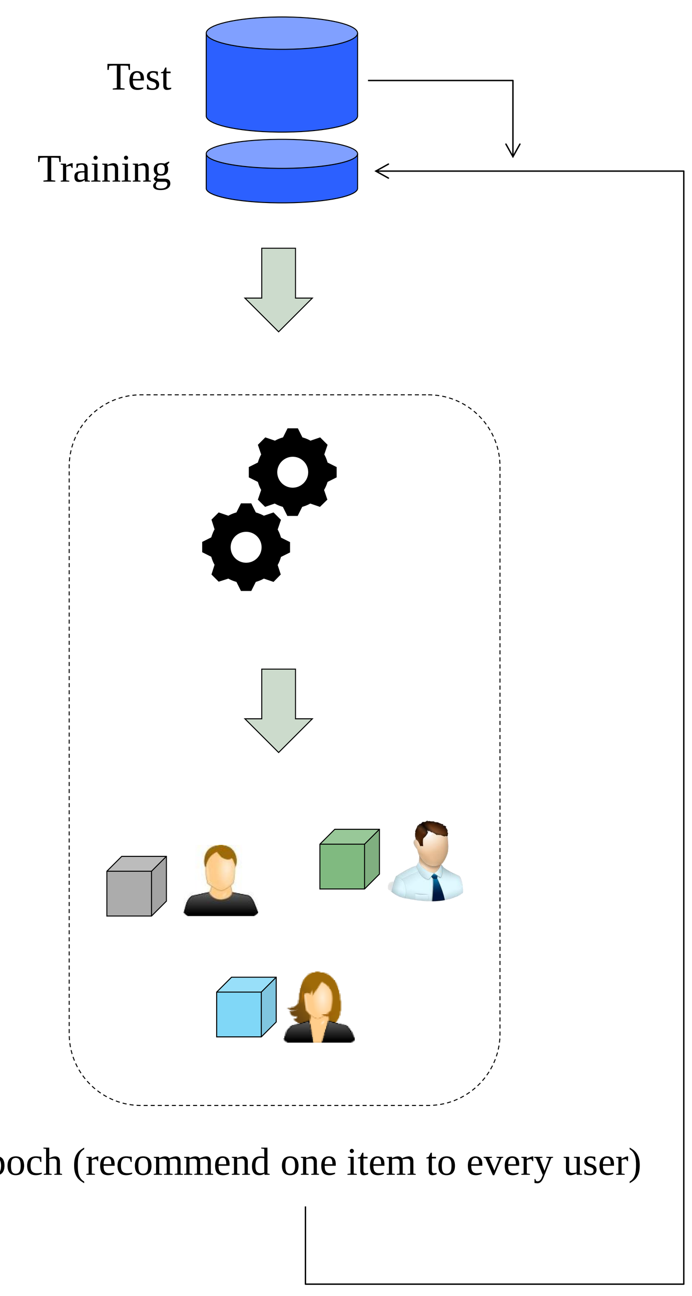
Reward ⟷ { 1 if the user likes the recommendation; 0 otherwise

Experimental Approach

Simulation loop

1. Initial random rating split

Each system has its own training and test sets (only the initial sets are common to all systems)



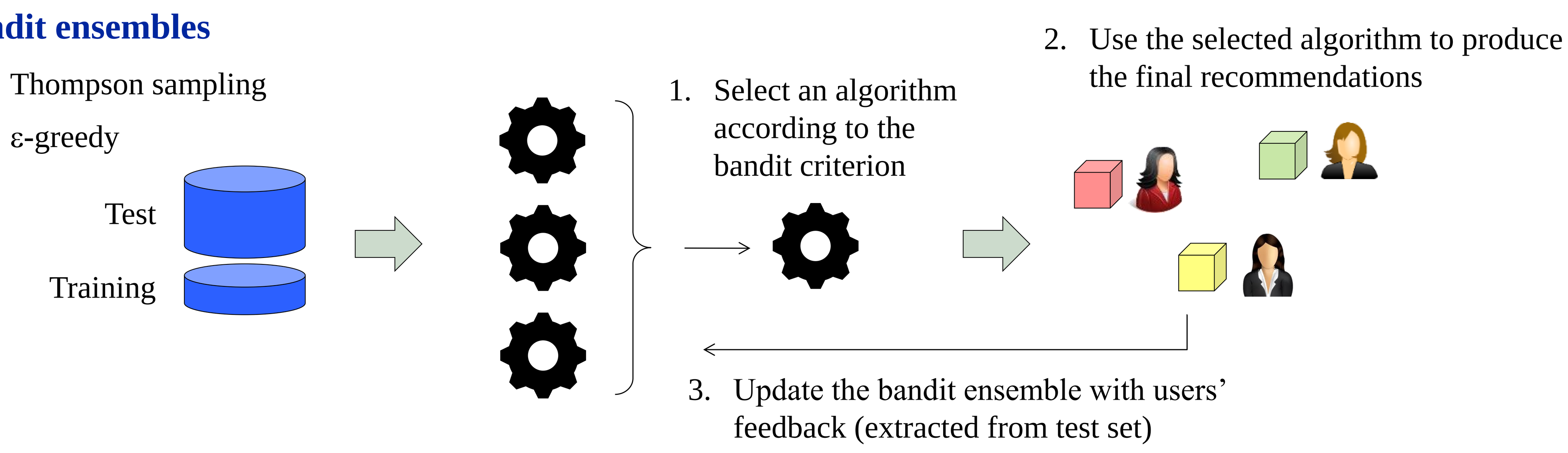
2. The system produces a recommendation for every user

3. After each epoch:
   - Compute cumulative recall with test set
   - Add users' feedback to training set and go back to step 2

Three types of systems: bandit ensembles, standalone algorithms and a dynamic non-bandit ensemble

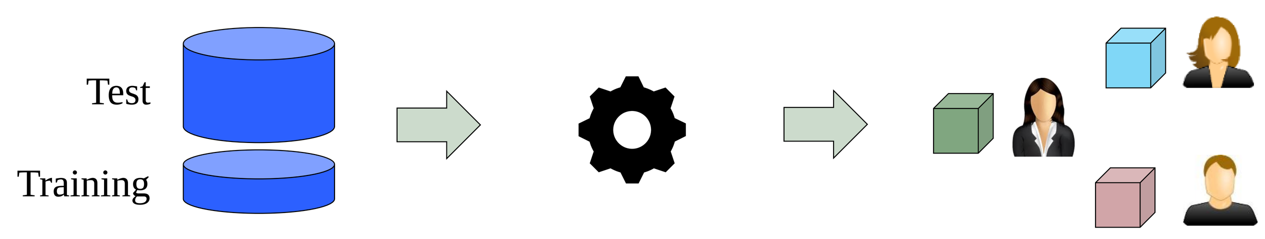
Bandit ensembles

- Thompson sampling
- ε-greedy



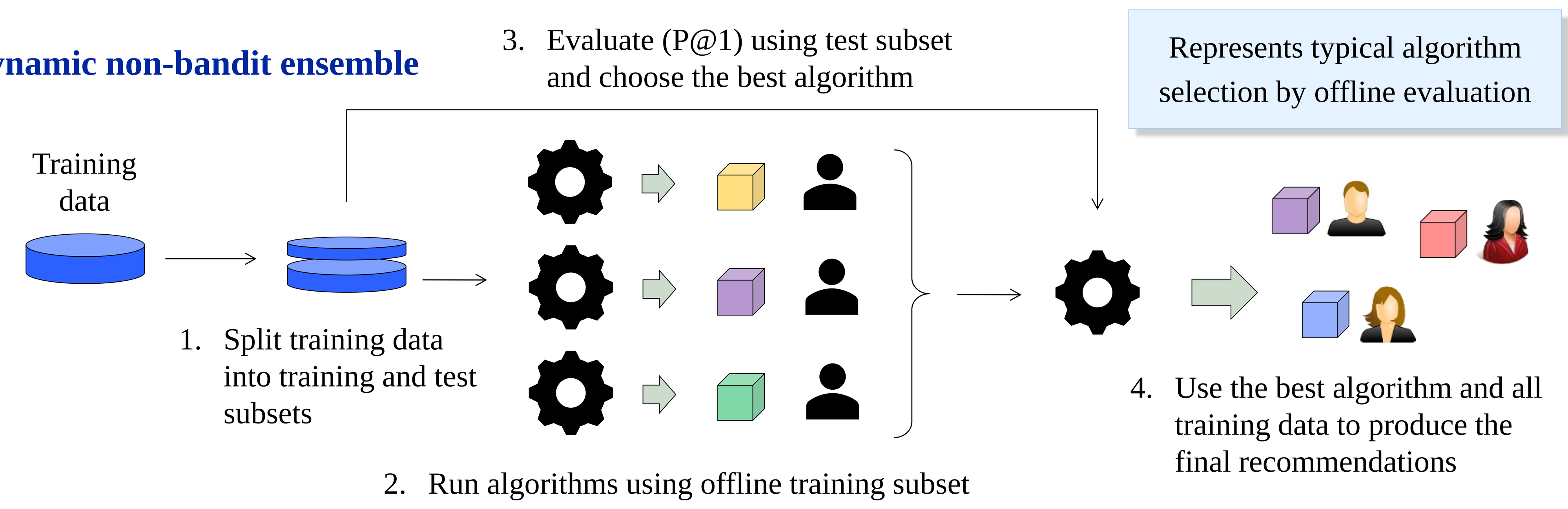
Standalone recommendation algorithms

- Matrix factorization
- User-based kNN
- Most popular



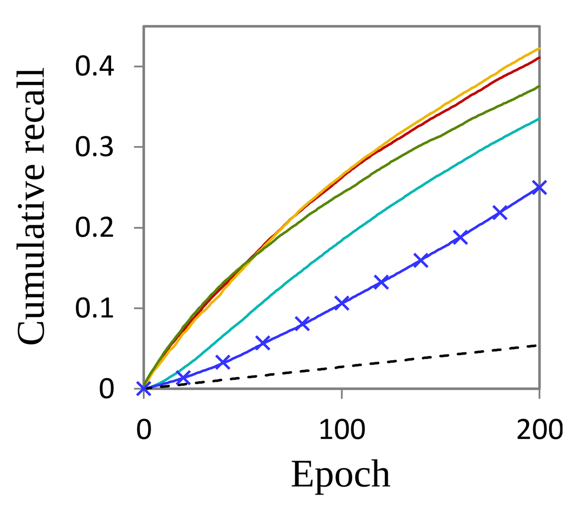
Dynamic non-bandit ensemble

Represents typical algorithm selection by offline evaluation



Results

Cumulative recall curves

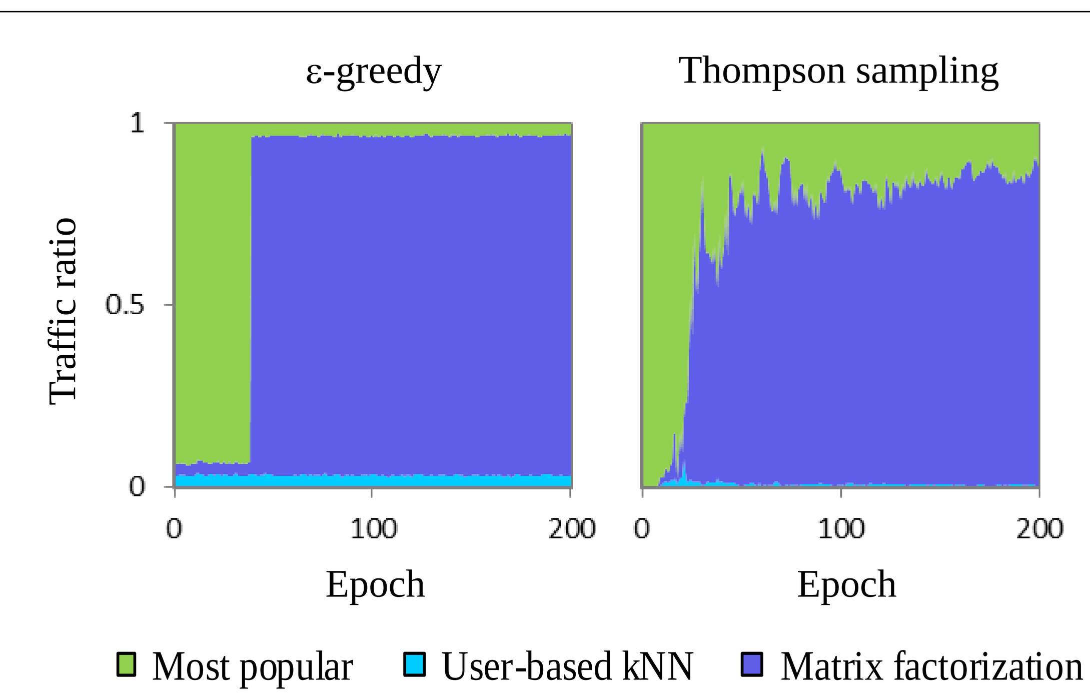


- Dataset: MovieLens 1M, 5% training / 95% test
- Bandit ensembles outperform all other systems
- Popularity > MF and kNN → Probably due to initial data sparsity
- Dynamic ensemble is stalled with matrix factorization → *Feedback loop*

Reduce initial split ratio to 1% and repeat experiment 4 times

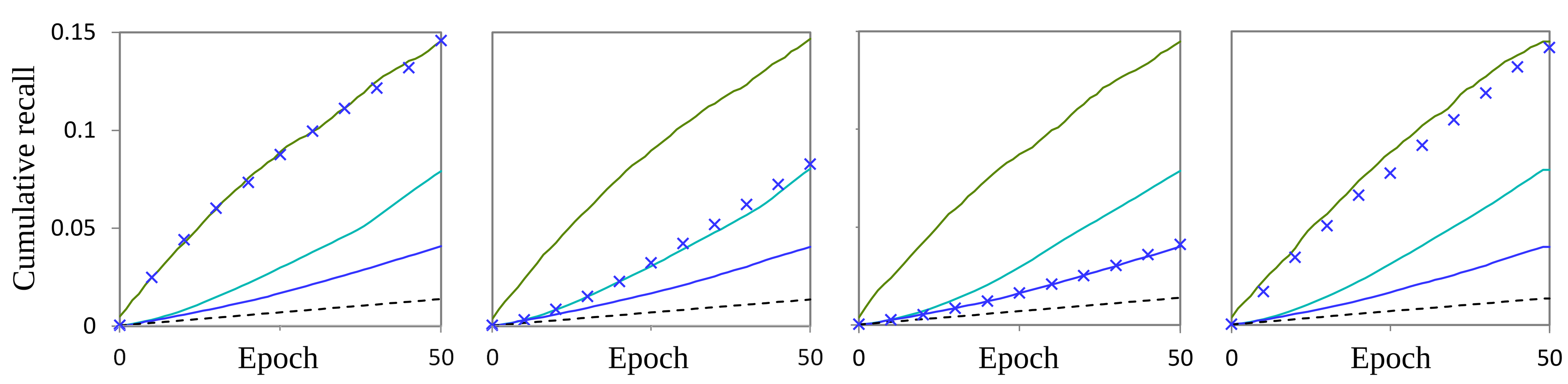
Traffic ratio of bandit ensembles

- Popularity dominates first iterations. then bandits turn to MF
- Not a clear winner → bandit ensemble as "infinite A/B testing"
- Low computational cost compared to dynamic ensembles: only one of the combined algorithms is run per epoch



Feedback loop in the dynamic non-bandit ensemble

- Thompson sampling bandit
- ε-greedy bandit
- Most popular
- User-based kNN
- Matrix factorization (MF)
- × Dynamic ensemble
- Random recommendation



- Variance of initial training set → different winners in first epoch
- In most cases, the algorithm selected in the first epoch stays the winner in the following iterations
- The ensemble fails to outperform the best algorithm most of the times

Code of the experiments is available at https://github.com/ir-uam/EnsembleBandits

Paper pdf

References

[1] B. Brodén, M. Hammar, B. Nilson and D. Paraschakis. Ensemble Recommendations via Thompson Sampling: an Experimental Study within e-Commerce (IUI 2018). Tokyo, Japan, 19–29.
[2] O. Chapelle and L. Li. An empirical evaluation of Thompson Sampling (NIPS 2011). Curran Associates, Inc., Red Hook, NY, USA, 2249–2257.
[3] D. Hill, H. Nassif, Y. Liu, A. Iyer and S. Vishwanathan. An Efficient Bandit Algorithm for Realtime Multivariate Optimization (SIGKDD 2017). Halifax, NS, Canada, 1813–1821.
[4] L. Li, W. Chu, J. Langford and R. Schapire. A contextual-bandit approach to personalized news article recommendation (WWW 2010). ACM, New York, NY, USA, 661–670.
[5] K. Pang, M. Dong, Y. Wu and T. Hospedales. Dynamic Ensemble Active Learning: A Non-Stationary Bandit with Expert Advice (ICPR 2018). IEEE Computer Society, Washington, DC, USA, 2269–2276.
[6] R. Sutton and A. Barto (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press, Cambridge, MA, USA, 2018.
[7] L. Tang, Y. Jiang, L. Li and T. Li. Ensemble contextual bandits for personalized recommendation (RecSys 2014). Foster City, CA, USA, 73–80.
[8] Q. Wang, C. Zeng, W. Zhou, T. Li, S. S. Iyengar, L. Shwartz and G. Grabarnik. Online Interactive Collaborative Filtering Using Multi-Armed Bandit with Dependent Arms (IEEE TKDE 2019), 31 (8), 1569–1580.

13th ACM Conference on Recommender Systems (RecSys 2019)