Infinite Recommendation Networks

A Data-Centric Approach

Question: Is *more data* what you need for *better recommendation*?

Noveen Sachdeva, Mehak Preet Dhaliwal, Carole-Jean Wu, Julian McAuley

UC San Diego   Meta

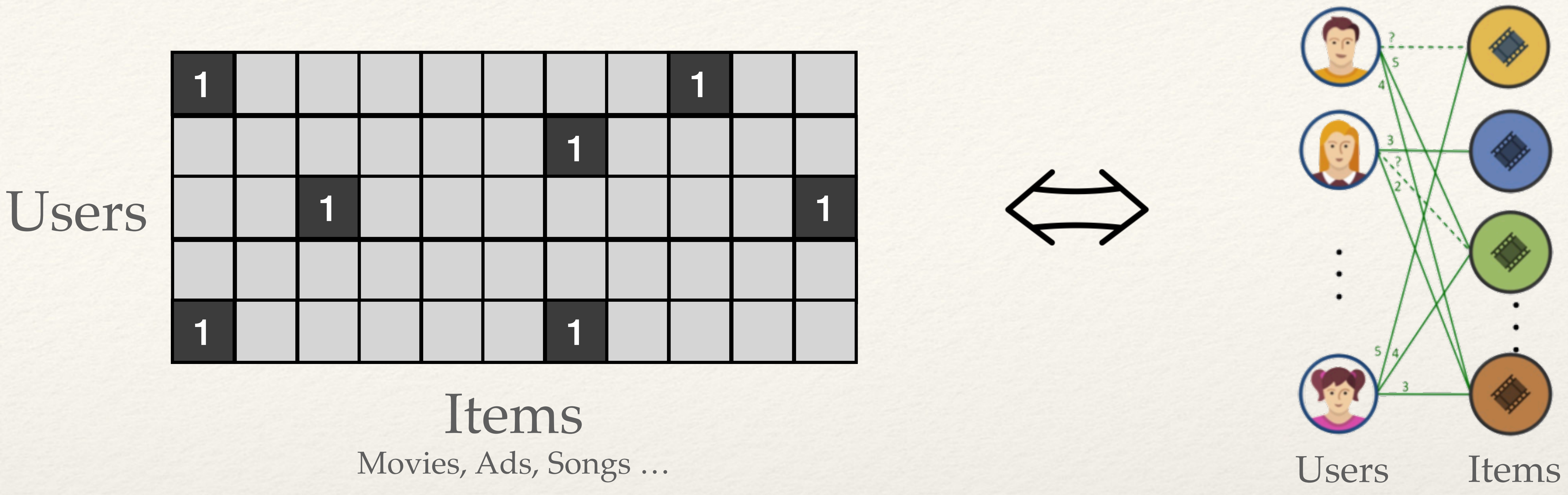
Data Perf Workshop @ ICML, July 2022

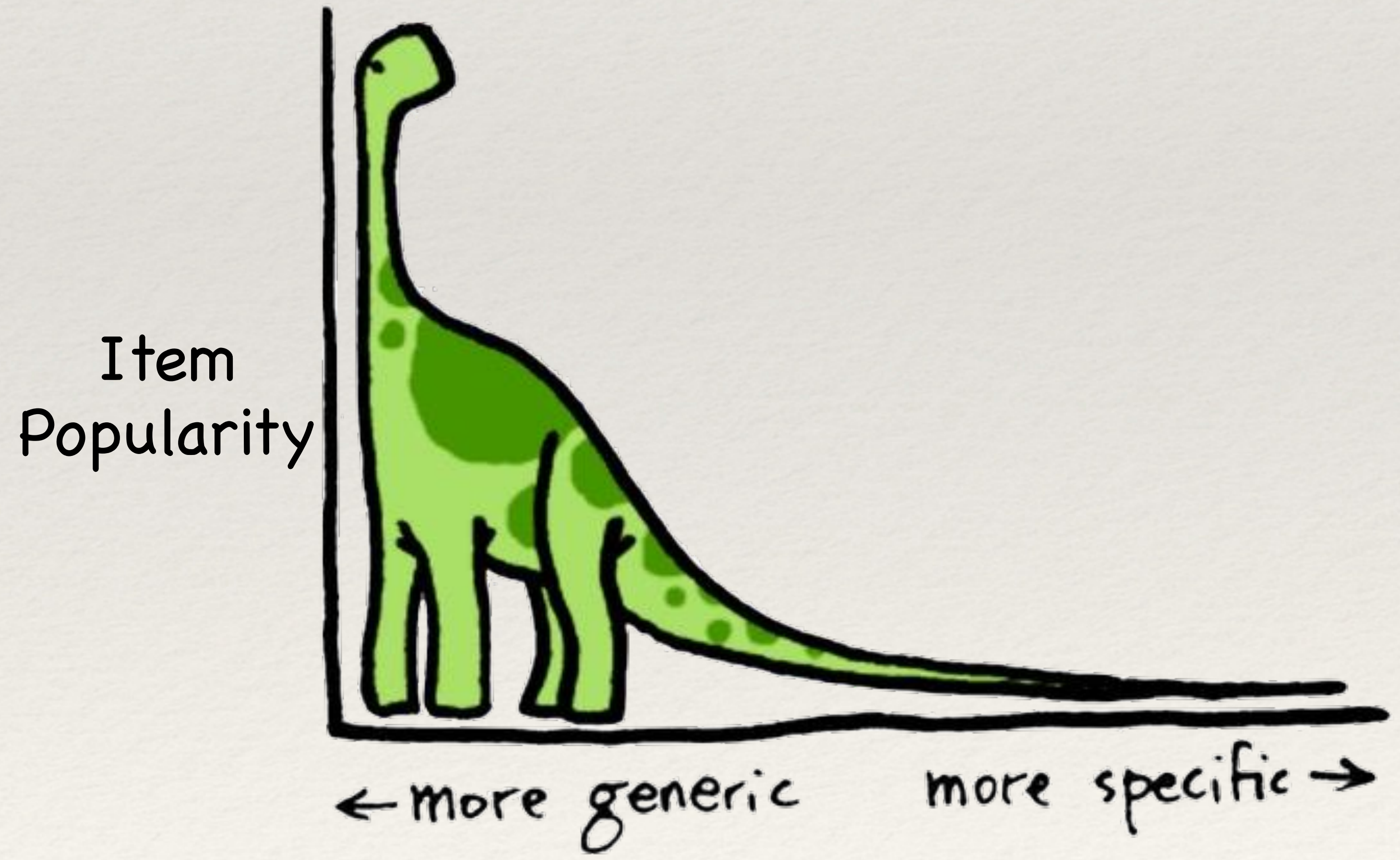
Research Objective

How to **synthesize** a small, representative summary of a collaborative filtering (CF) dataset which can accurately retain the **performance** of algorithms trained on the full dataset *vs.* on the data summary?

Challenges:

- Data heterogeneity
- Semi-structuredness
- Sparsity & Long-tail characteristics

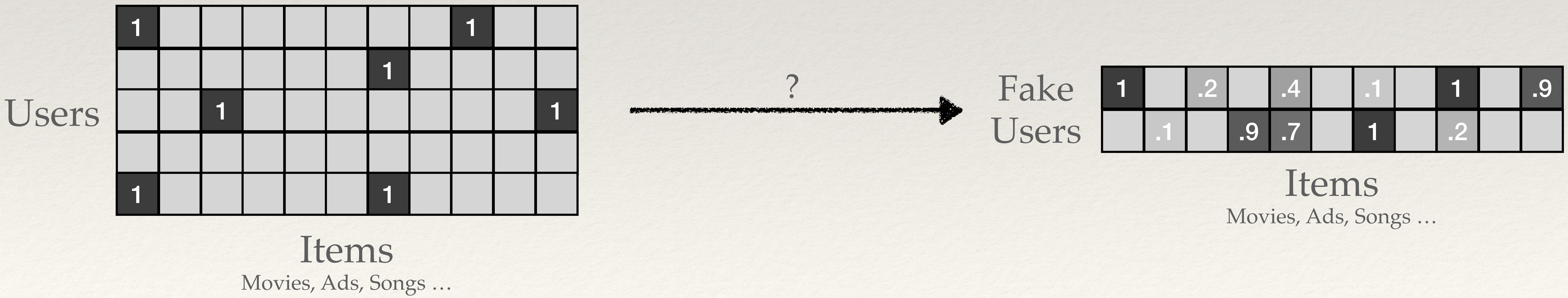




Distill-CF

Data Distillation for Collaborative Filtering Data

<u>Premise</u>: Treat the to-be-synthesized data as **parameters**, and **learn** them through a bilevel optimization.



Distill-CF

Data Distillation for Collaborative Filtering Data

Robust framework:

- Uses Gumbel sampling on $\mathcal{D}^s$ to mitigate the heterogeneity of the problem
- Perform Gumbel sampling multiple times for each fake-user to handle dynamic user/item popularity
- **Optimizes** for data-quality rather than quantity

Outer loop — optimize the support set for a fixed learning algorithm

Support dataset

Differentiable cost-function

$$\underset{\boxed{\mathcal{D}^s}}{\arg\min} \underset{(u,i,r)\sim\mathcal{D}^{\text{val}}}{\mathbb{E}} \left[\boxed{l}\left(\phi^*_{\mathcal{D}^s}(u,i),r\right)\right]$$

$$\text{s.t.} \quad \boxed{\phi^*_{\mathcal{D}^s}} := \underset{\phi}{\arg\min} \underset{(u,i,r)\sim\mathcal{D}^s}{\mathbb{E}} \left[l\left(\phi(u,i),r\right)\right]$$

Optimal recommendation algorithm trained on $\mathcal{D}^s$

Inner loop — optimize the learning algorithm for a fixed support set

Experiments

Major Results

- Using Distill-CF, we can get **96-105%** of full-data performance on as small as **0.1%** data sub-samples, leading to as much as **~1000x** time speedup!

- Distill-CF is robust to noise (even though not optimized for it), and is able to offer significant performance at high noise ratios, even with very small support datasets!

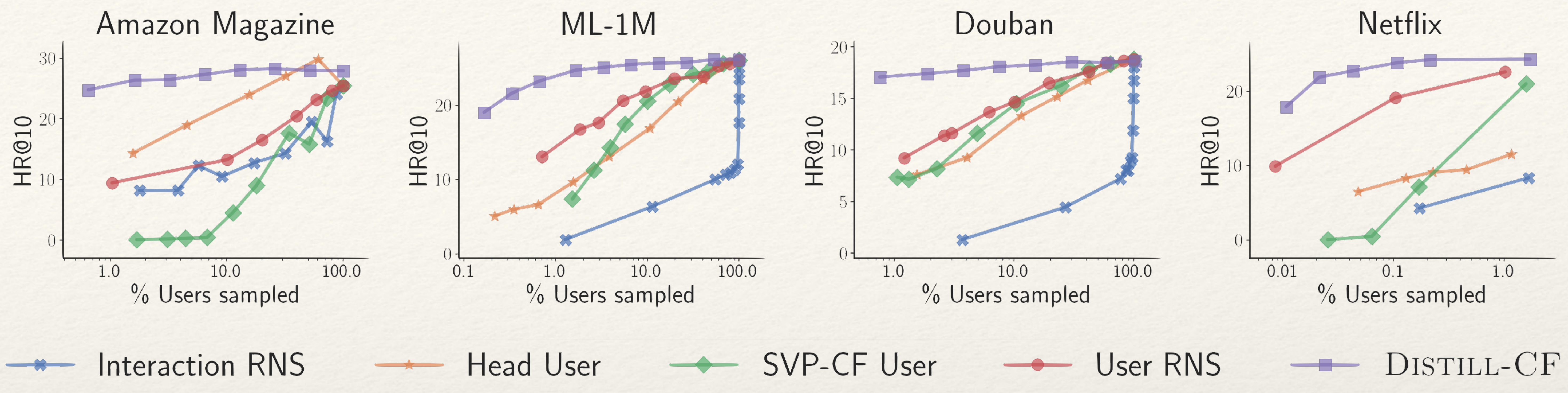


Figure 1: Does Distill-CF outperform other samplers? (Log-scale)

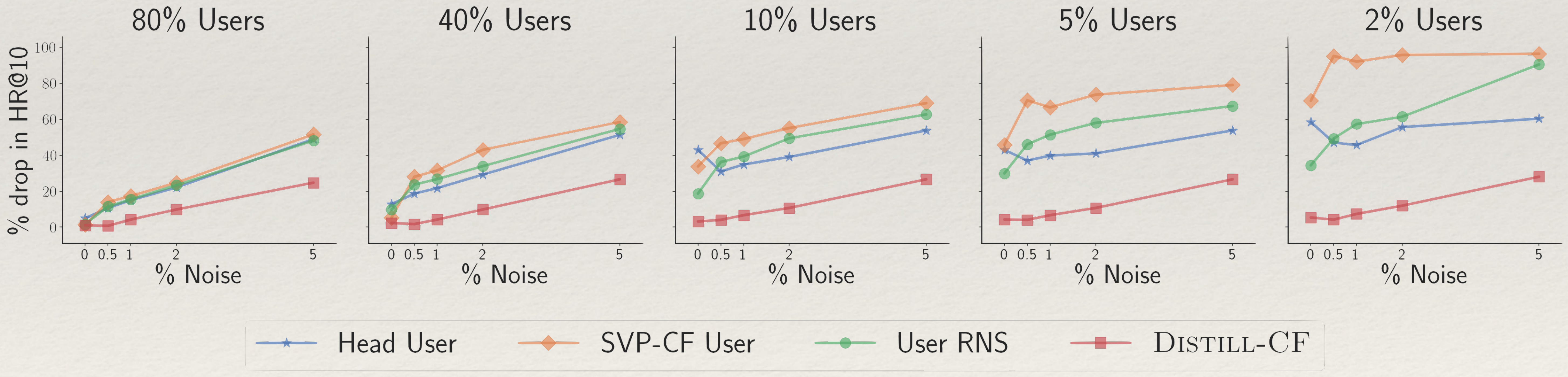


Figure 2: How does noise in the data affect different samplers?

Thank you!
Questions?

@noveens97

For paper, code, and these slides: