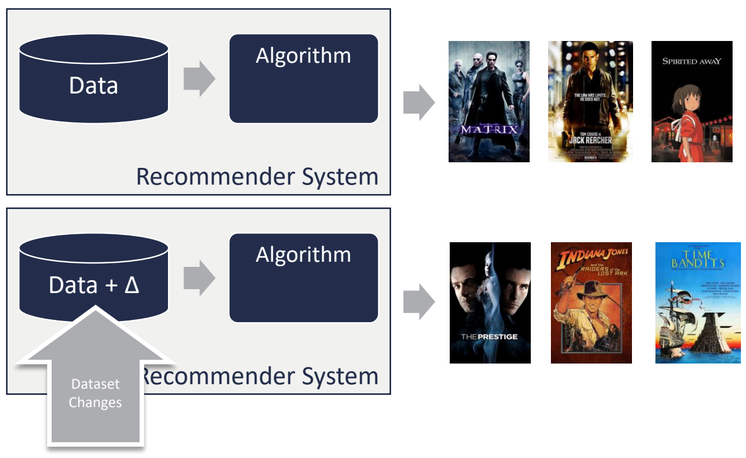
Evaluating Recommender System Stability with Influence-Guided Fuzzing

David Shriver[1] Sebastian Elbaum[1] Matthew B. Dwyer[1] David S. Rosenblum[2]

[1]University of Virginia [2]National University of Singapore

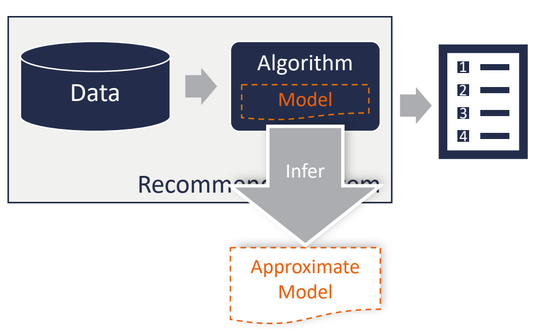
Problem



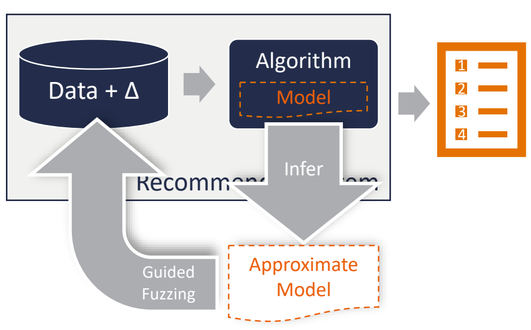
Small modifications to a dataset can cause drastic recommendation changes, negatively impacting users and businesses.

How can we cost-effectively find small dataset modifications that induce such drastic recommendation changes?

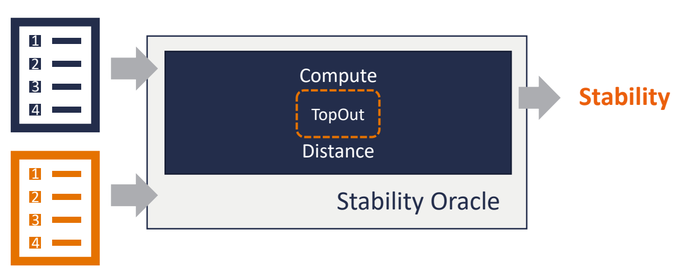
Conceptual Solution



1. Infer approximate model of the implicit relationships learned by the recommendation algorithm.



2. Leverage inferred model to generate dataset modifications that are more likely to induce instability.



3. Compute the distance between recommendations before and after dataset modifications to assess recommender stability.

Influence-Guided Fuzzing

Influence Models

User | Item | Attribute | Rating

$$I^U(u) = |\{u' | rank(i, u') \neq \perp \wedge rating(u, i)\}|$$

A user that rates items recommended to other users is influential to the system.

Modification Types

Add a rating to the dataset.

Remove a rating from the dataset.

Change a rating in the dataset.

Sample Heuristics

Add a rating with random value to a random item for the *most* influential **user**.

Remove a rating from a random item for the *most* influential **user**.

Add a rating for a random user, with a random value, to the *least* influential **item.**

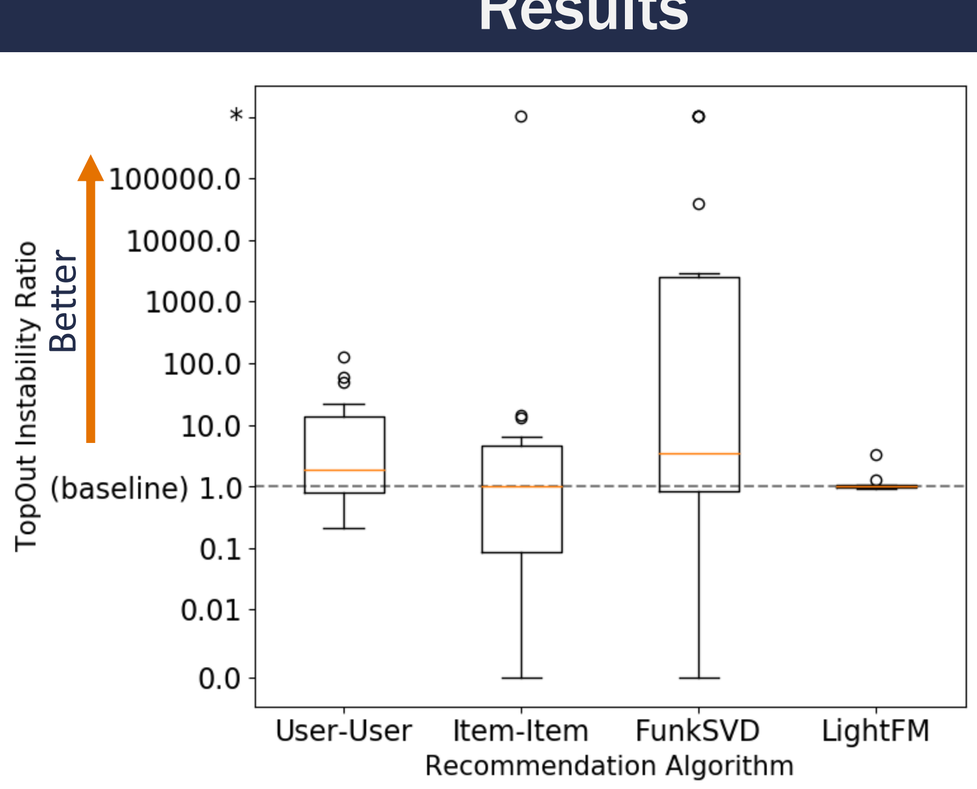
Results

4 Recommender Algorithms
User-User, Item-Item, FunkSVD, LightFM

3 Baseline Heuristics

MovieLens dataset
100,000 ratings

3 Modification Set Sizes
1, 10, 100



Conclusions

1. Influence-guided fuzzing is more effective than random fuzzing, especially when the influence model matches underlying recommendation algorithms.

2. Using a portfolio approach is effective for fuzzing modification sets that induce significant change.

This work has been supported in part by National Science Foundation awards #1526652 and #1617916, and by A*STAR SERC PSF grant 152120008.