Personalized Reward Learning with Interaction-Grounded Learning (IGL)

Jessica Maghakian, Paul Mineiro, Kishan Panaganti, Mark Rucker, Akanksha Saran, Cheng Tan

Data sparsity is challenging!

Explicit user feedback is rare in recommender systems. As a result, state-of-the-art (SOTA) systems use implicit signals instead:

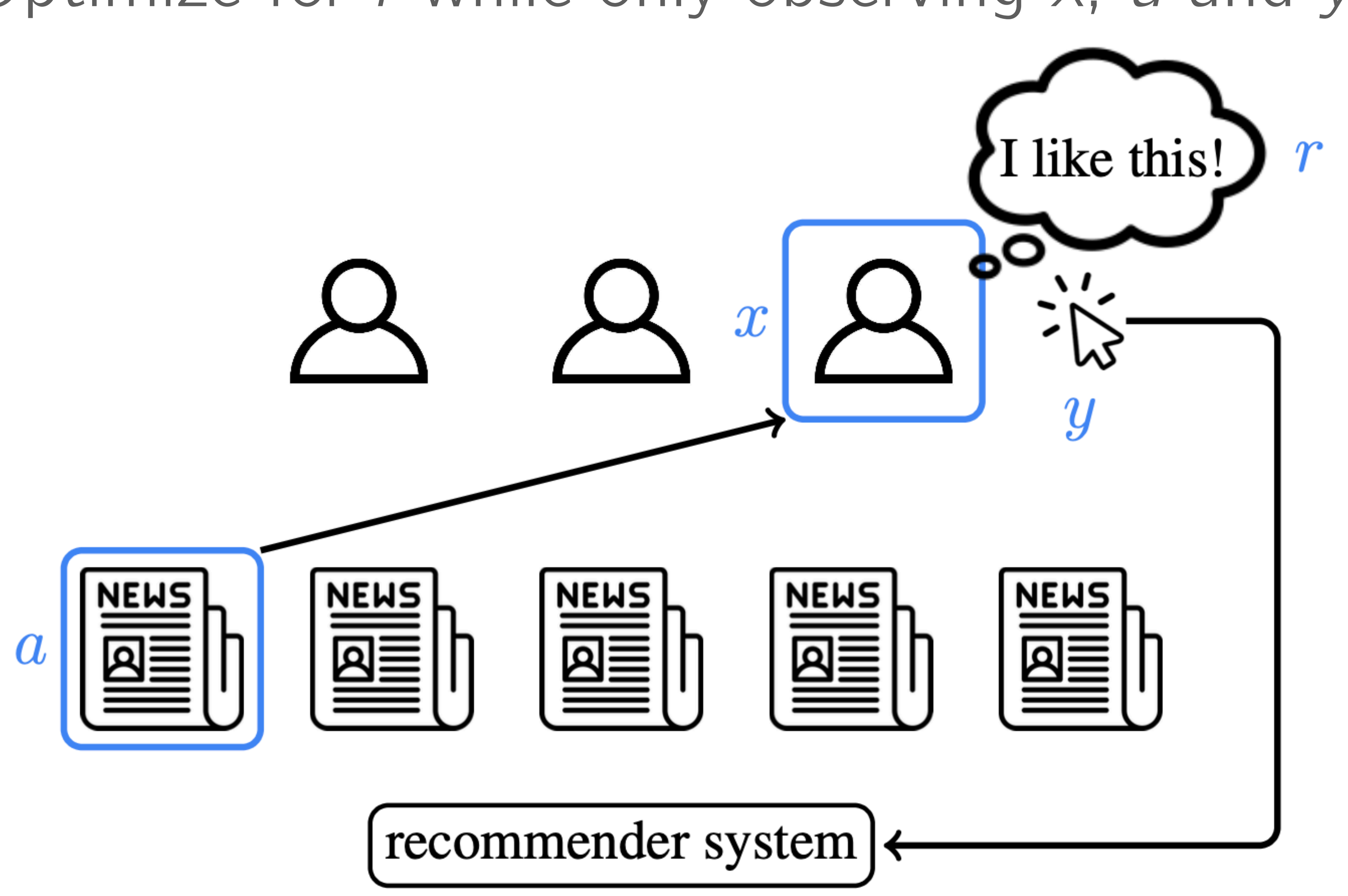
- Facebook in 2016:
  r = **1**👍 + **5**❤️ + **5**😆 + **5**😮 + **5**😢 + **5**😡
- Twitter in 2023:
  r = **27**💬 + **1**🔁 + **0.5**♡

But these weighted combinations are not the true reward and have many limitations!

Our idea: directly maximize latent reward using Interaction-Grounded Learning (IGL)

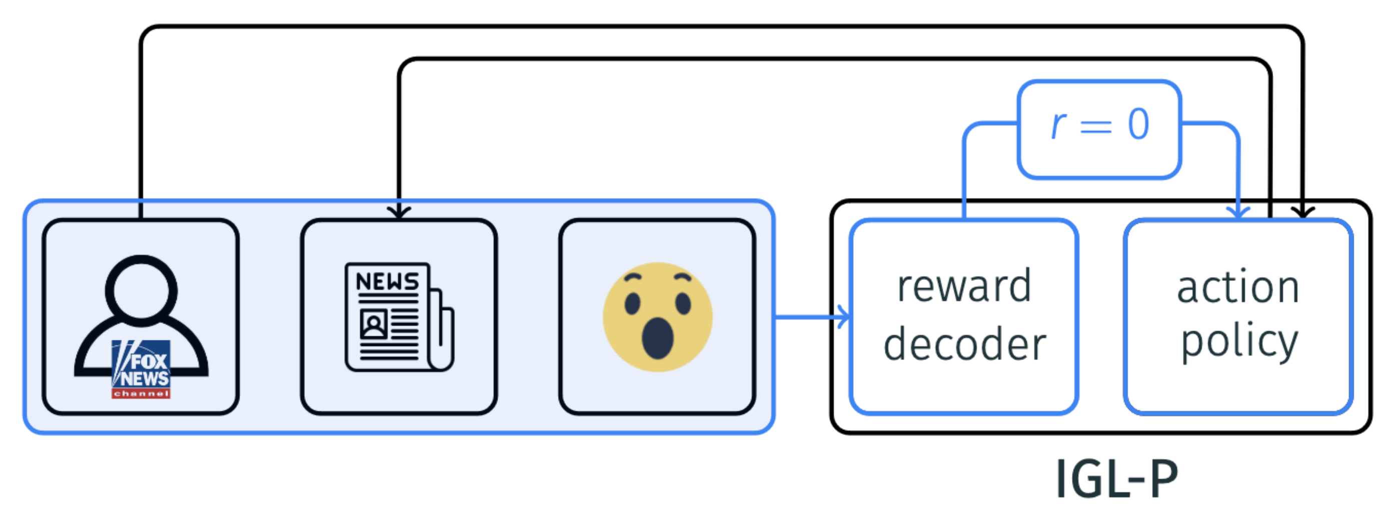
What is IGL for recommender systems?

Optimize for r while only observing x, a and y



Our solution: IGL-P for personalization

IGL-P needs just 2 conditions to succeed: rare rewards and consistent communication



IGL-P efficiently learns different reward functions for different users

Result: IGL-P matches production policy

We first evaluated IGL-P using millions of interactions from production data of image recommendation for Windows users

Clicks	[1.000, 1.010, 1.020]
Likes	[1.006, 1.026, 1.049]
Dislikes	[0.890, 0.918, 0.955]

Table 1. Relative metrics lift of IGL-P over production policy (point estimate and 95% CI)

IGL-P can match SOTA hand-engineered baseline at a fraction of the cost!

Result: IGL-P improves user fairness

Using 2016 Facebook news data we evaluated and compared IGL-P to contextual bandit policies trained with two reward mappings previously used by Facebook.

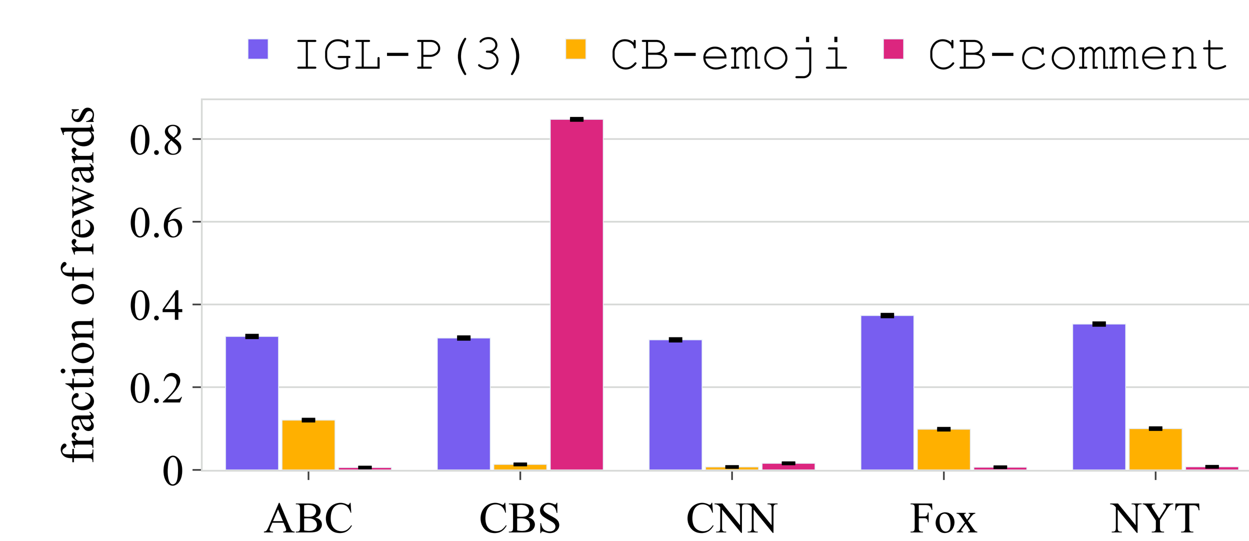


Figure 1. Average fraction of positive rewards (with standard error) across different user types

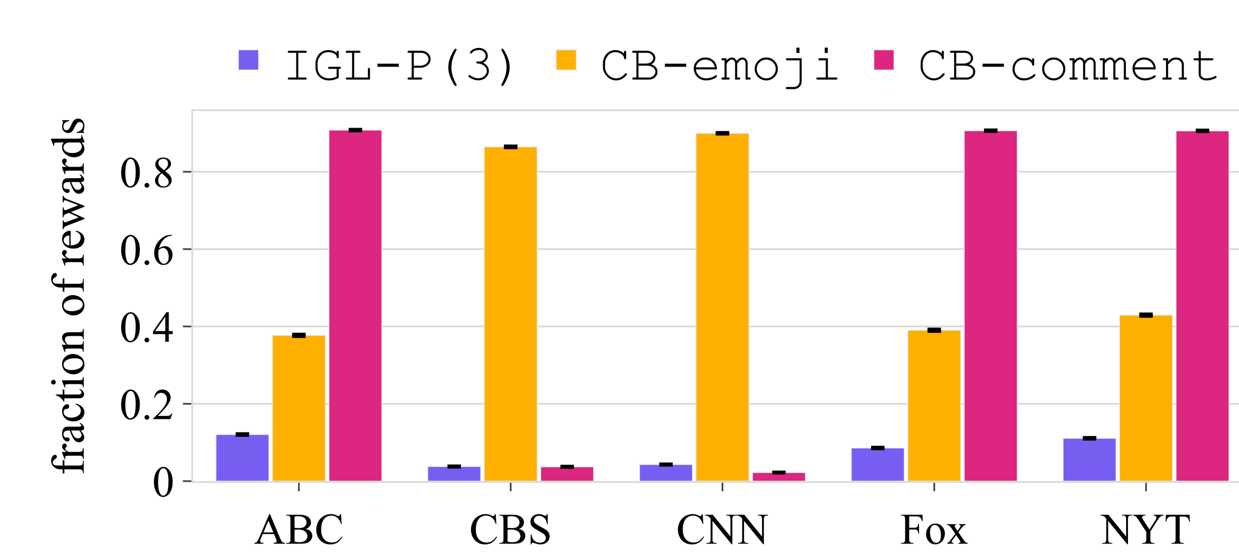


Figure 2. Average fraction of negative rewards (with standard error) across different user types

IGL-P beyond recommender systems

Do you think personalized reward learning can benefit your application? Send me an email at: `jessica.maghakian@stonybrook.edu`