Personalized Reward Learning with Interaction-Grounded Learning (IGL)

Jessica Maghakian[1], Paul Mineiro[2], Kishan Panaganti[3], Mark Rucker[4], Akanksha Saran[2], Cheng Tan[2]

[1]Stony Brook University, [2]Microsoft Research NYC, [3]Texas A&M University, [4]University of Virginia

ICLR 2023
Kigali, Rwanda
May 1 – May 5, 2023

Reward engineering for recommender systems is complicated

Goal: show users that content they like and enjoy

Reward engineering for recommender systems is complicated

Goal: show users that content they like and enjoy

Challenge: explicit user feedback is rare in recommender systems

Reward engineering for recommender systems is complicated

Goal: show users that content they like and enjoy

Challenge: explicit user feedback is rare in recommender systems

SOTA: find a “good” weighted combination of implicit feedback

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

Reward engineering for recommender systems is complicated

Goal: show users that content they like and enjoy

Challenge: explicit user feedback is rare in recommender systems

SOTA: find a "good" weighted combination of implicit feedback

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

Using fixed weighting of implicit feedback is not ideal...

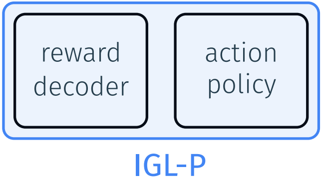
- weights can be arbitrary with unanticipated consequences
- implicit signals are nuanced and complicated
- weights require continuous updating as users and UI evolve
- can result in unfair one-size-fits-all systems

Personalized reward learning with IGL-P

Idea: *learn* personalized reward functions through user interactions

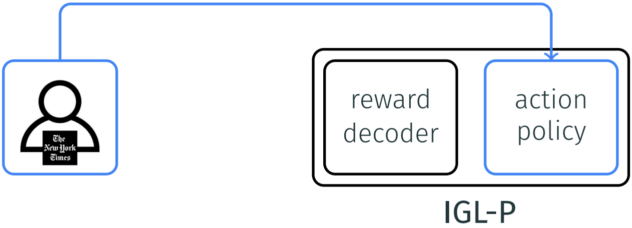
Personalized reward learning with IGL-P

Idea: *learn* personalized reward functions through user interactions



Personalized reward learning with IGL-P

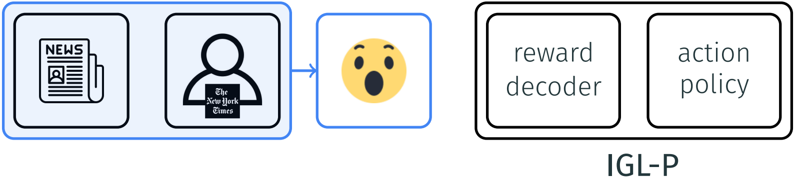
Idea: *learn* personalized reward functions through user interactions



Idea: *learn* personalized reward functions through user interactions

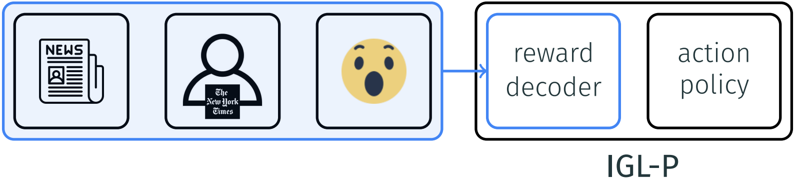
Personalized reward learning with IGL-P

Idea: *learn* personalized reward functions through user interactions



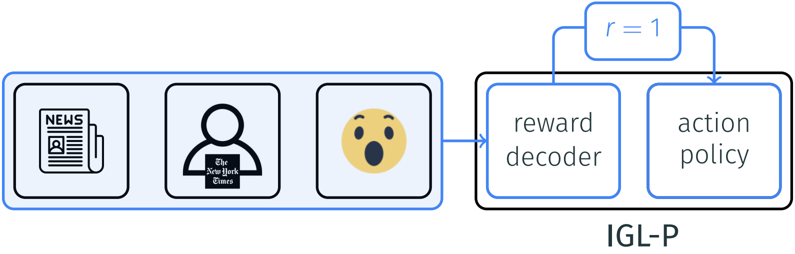
Personalized reward learning with IGL-P

Idea: *learn* personalized reward functions through user interactions



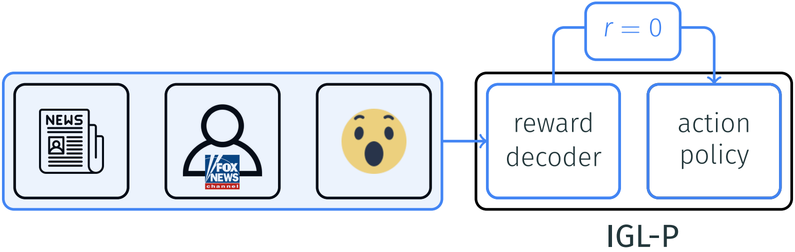
Personalized reward learning with IGL-P

Idea: *learn* personalized reward functions through user interactions



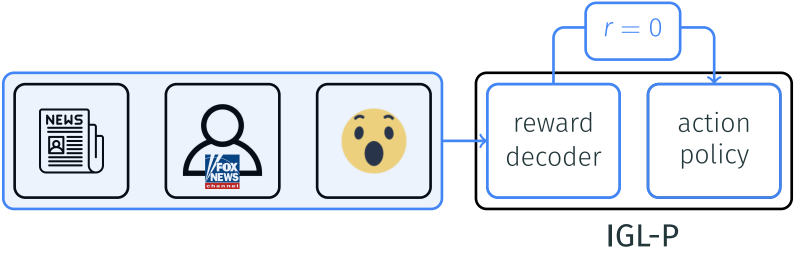
Personalized reward learning with IGL-P

Idea: *learn* personalized reward functions through user interactions



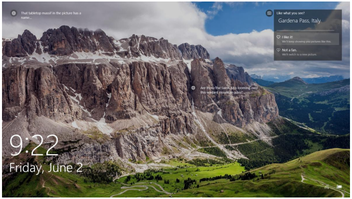
Personalized reward learning with IGL-P

Idea: *learn* personalized reward functions through user interactions



IGL-P only requires two simple conditions to succeed:
(1) rewards are rare and (2) users communicate consistently

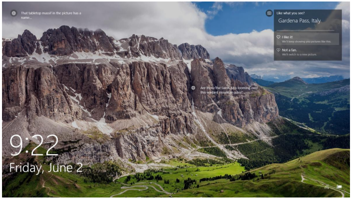
IGL-P SUCCEEDS ON REAL WORLD DATA



Exp. 1: Image recommendation for Windows users

- IGL-P matched the state-of-the-art production policy that was trained on significantly more data for positive feedback signals
- IGL-P outperformed production policy with respect to negative signals

IGL-P SUCCEEDS ON REAL WORLD DATA



Exp. 1: Image recommendation for Windows users

Exp. 2: News recommendation for Facebook users

- IGL-P matched the state-of-the-art production policy that was trained on significantly more data for positive feedback signals
- IGL-P outperformed production policy with respect to negative signals

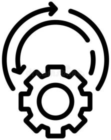
- Competitor policies trained with rewards used by Facebook circa 2017 offered unfair performance across different user types
- IGL-P performed consistently well across different user types

Summary of our results

IGL-P can match state-of-the-art performance at a fraction of the cost

Summary of our results

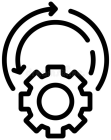
IGL-P can match
state-of-the-art
performance
at a fraction
of the cost



IGL-P can easily
adapt and evolve
with changing
systems
and users

Summary of our results

IGL-P can match state-of-the-art performance at a fraction of the cost

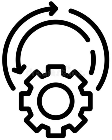


IGL-P can easily adapt and evolve with changing systems and users

IGL-P uses personalized rewards to improve fairness for diverse users

Summary of our results

IGL-P can match state-of-the-art performance at a fraction of the cost



IGL-P can easily adapt and evolve with changing systems and users

IGL-P uses personalized rewards to improve fairness for diverse users

Although we introduced personalized reward learning for recommender systems, IGL-P can benefit any application that suffers from a one-size-fits-all approach!