



Policy optimization:

- In complex NGN scenarios, learning from expert trajectories can maximize the efficiency of policy refinement.
- The reward learned from IRL can maximize the effectiveness for **mimicking expert trajectories** (E).

Motivations for applying IRL

Environmental complexity:

- The **network complexity and diversity** increase due to the introduction of new communication and networking paradigms, such as SAGIN, and 6G.
- Numerous physical factors** explicitly or implicitly affect environment, causing difficulty in reward modeling (D).

Reward unavailability:

- In many cases of NGN, the reward cannot be modeled, such as
- The network contains **adversarial**, e.g., attackers, eavesdroppers, whose objective is **hidden** from users (A).
- The immediate reward to each action is **unavailable** (B).
- The NGN is **human-in-the-loop** (C).

