

1. Tầm quan trọng của Model trong RL

- Sử dụng mô hình học được từ dữ liệu giúp ta:
 - Học trong mô phỏng (simulation).
 - Tiết kiệm chi phí và thời gian so với học trực tiếp trong môi trường thật.
-

2. Quá trình truyền thống: System Identification

- Dựa trên Supervised Learning:
 - Thu thập tập dữ liệu `state, action → next state`.
 - Học một mô hình động lực học (transition function) $T^{\wedge}(s,a) \rightarrow \hat{s'}(s, a) \rightarrow s'$.
 - Kết hợp với cost function + planner để sinh ra policy.
 - ❌ **Vấn đề:** mô hình được học có thể **không chính xác ở những vùng quan trọng**, do:
 - Thiếu dữ liệu ở vùng mà policy sẽ thường xuyên truy cập.
 - → **Exploit lỗi mô hình** → policy kém khi áp dụng thực tế.
-

3. Vấn đề cốt lõi: Covariate Shift

- Mô hình học từ dữ liệu **không đại diện cho toàn bộ vùng trạng thái** mà agent sẽ gặp.
 - Policy học được có thể **khai thác lỗi** trong mô hình → gây ra **hiệu suất thấp trong thế giới thực**.
-

4. Quan điểm mới: Học mô hình tương tác (Iterative Interactive Model Learning)

Chu trình tương tác lặp đi lặp lại:

1. **Khởi đầu:** thu thập dữ liệu từ **policy ban đầu** (exploration hoặc expert).
2. Học mô hình $T^{\hat{T}}$ từ dữ liệu đã có.
3. Dùng $T^{\hat{T}}$ và cost function để **sinh ra policy tối ưu mới**.
4. Dùng policy mới để **thu thập thêm dữ liệu** → lặp lại quá trình.
5. Dữ liệu **được tích lũy dần qua mỗi vòng lặp**.

Đây là ý tưởng của thuật toán như **DAGGER (Dataset Aggregation)**.

5. Lợi ích của cách học tương tác

- Giống phong cách làm việc thực tế của kỹ sư.
 - **Tăng tính ổn định**, mô hình cải thiện dần theo dữ liệu thật.
 - Có **bảo đảm lý thuyết** (theoretical guarantees):
 - Nếu dùng **hàm xấp xỉ ổn định** và thuật toán tối ưu tốt → đạt hiệu suất tốt.
 - Ngược lại: nếu kết quả tệ → chắc chắn là do **mô hình chưa đủ chính xác**.
-

6. No-Regret Learning

- Khái niệm mạnh mẽ dùng trong nhiều bối cảnh giống game.
- Đặc điểm:
 - Mô hình không thay đổi mạnh khi dữ liệu huấn luyện thay đổi nhỏ.
 - Hiệu suất tiệm cận tối ưu khi có nhiều dữ liệu.

