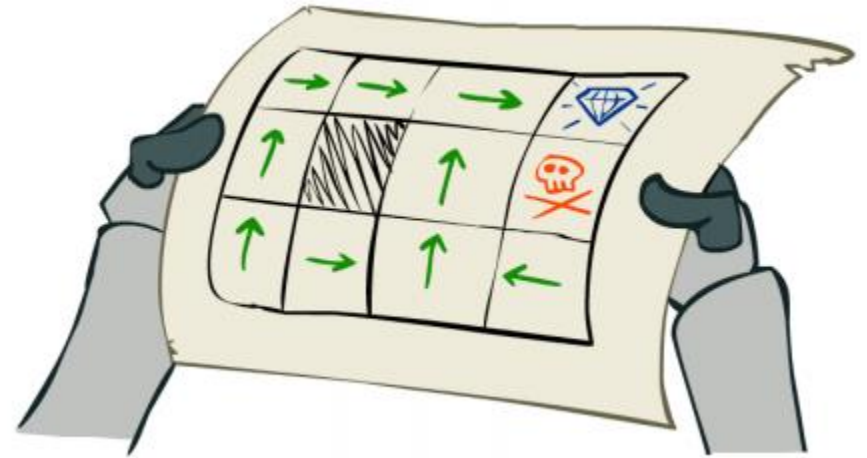
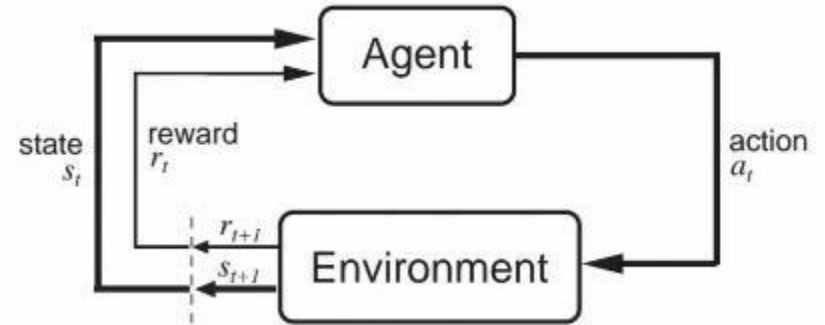


마르코브 의사결정 프로세스 예시



MDP 예시 I

• 주사위 던지기 게임

- 매 라운드 마다, 당신은 go 혹은 stop 을 할 수 있다
- 만약 stop를 선택하면 \$8을 받고 게임 종료
- 만약 go를 선택하면, \$4을 받고 주사위를 돌린다
 - 만약 주사위 결과가 1 혹은 2라면, 게임 종료
 - 아니면 다음 라운드로 진행

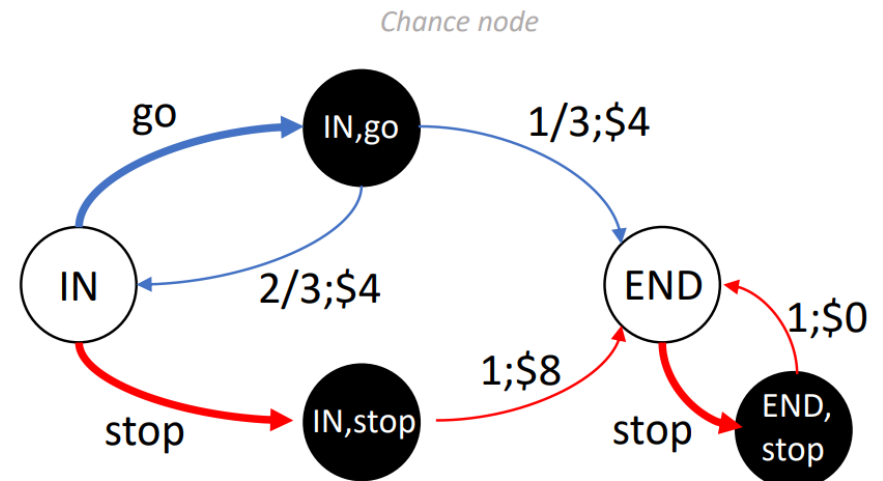
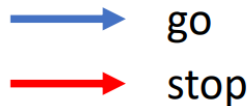
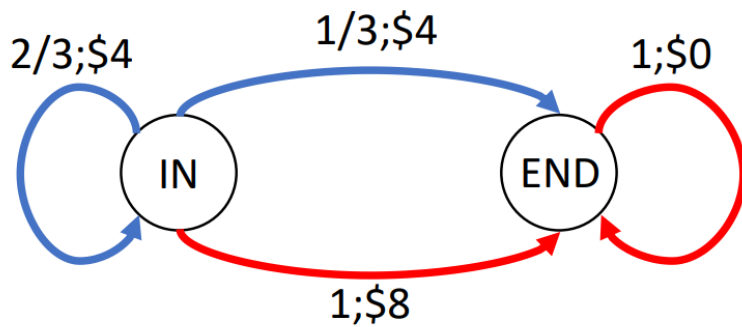
• MDP 모델링

- 의사결정 시점: $T = \{1, 2, 3, \dots\}$
- 상태공간: $S = \{in, end\}$
- 행동공간: $A_{in} = \{go, stop\}, A_{end} = \{stop\}$
- 상태전이확률
 - $P_t(in|in, go) = 2/3$
 - $P_t(end|in, go) = 1/3$
 - $P_t(in|in, end) = 0$
 - $P_t(end|in, end) = 1$
 - $P_t(in|end, stop) = 0$
 - $P_t(end|end, stop) = 1$
- 보상
 - $r_t(in, stop) = \$8, r_t(in, go) = \$4, r_t(end, stop) = 0$

MDP 예시 I

• 주사위 던지기 게임

- 매 라운드마다, 당신은 go 혹은 stop 을 할 수 있다
- 만약 stop를 선택하면 \$8을 받고 게임 종료
- 만약 go를 선택하면, \$4을 받고 주사위를 돌린다
 - 만약 주사위 결과가 1 혹은 2라면, 게임 종료
 - 아니면 다음 라운드로 진행



MDP 예시2

• 재고관리 문제

• 문제 설명 및 가정

- 주문은 매달 시작 시점에 주문하고 주문량이 도착
- 수요(주문요청)는 한달 간 이루어지며, 주문은 매달 마지막 날에 일괄 충족
- 만약 월간 총 수요 > 재고, 미충족 수요는 상실 (품절처리)
- 수익 및 비용함수, 수요분포는 시간에 따라 변하지 않음
- 물류센터는 최대 M 개의 물품을 보관할 수 있음

• 기호

- s_t : t 번째 달 시작시점에서의 재고수준 (주문 전)
- a_t : t 번째 달 시작시점에서의 주문량
- D_t : t 번째 달 수요 ($p_j = P(D_t = j)$)
- $F(u)$: 재고수준이 u 일때의 기대 수익
- $o(u)$: 주문량이 u 일때의 주문 및 구매 비용
- $h(u)$: 재고수준이 u 일때의 한달동안의 재고보유비용

MDP 예시2

• 재고관리 문제

- 의사결정 시점: $T = \{1, 2, 3, \dots\}$
- 상태공간: $S = \{0, 1, 2, \dots, M\}$
- 행동공간: $A_s = \{0, 1, 2, \dots, M - s\}$
- 상태전이확률

$$p_t(j|s, a) = \begin{cases} 0 & \text{if } M \geq j > s + a \\ p_{s+a-j} & \text{if } M \geq s + a \geq j > 0 \\ \sum_{j:j \geq s+a} p_j & \text{if } M \geq s + a, j = 0 \end{cases}$$

• 보상

- $r_t(s, a) = F(s + a) - O(a) - h(s + a)$

• 정책 (혹은 의사결정규칙) 예시

$$\delta_t(u) = \begin{cases} 0 & \text{if } u \geq s \\ S - u & \text{if } u < s \end{cases}$$