MDP : 비결정적 탐색 문제, Offline Planning

- Expectimax로 해결 가능 : 전체 정책을 결정X
- Stationary Preference : $a > b$면 앞에 하나를 추가해도 $a > b$
- Finite Horizon : 일정 깊이에 도달 시 게임 종결
- Discounting : 무한등비급수에 의해 수렴
- Absorbing State : 언젠가는 종료 상태에 도달

벨만 방정식

- $V^*(s) = \max_a Q^*(s,a)$
- $Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$
- $V^*(s) = \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$
- k=0일 때 모든 상태의 값을 0으로

Value Iteration : 벨만 방정식의 계산 방법

- $V_{k+1}(s) \leftarrow \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V_k(s')]$
- 각 반복의 복잡도는 $O(S^2A)$, 단일 최적값으로 수렴
- V_k와 V_{k+1}의 차이는 최대 $\gamma^k \max|R|$이므로 수렴

Policy Evaluation : 정책을 고정하고 값을 계산

- $V_0^\pi(s) = 0$
- $V_{k+1}^\pi(s) \leftarrow \sum_{s'} T(s,\pi(s),s')[R(s,\pi(s),s') + \gamma V_k^\pi(s')]$

Policy Extraction : 최적값에서 정책 추출

- $\pi^*(s) = \arg\max_a Q^*(s,a)$
- $\pi^*(s) = \arg\max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$
- 행동을 선택하려면 V보단 Q가 더 유리

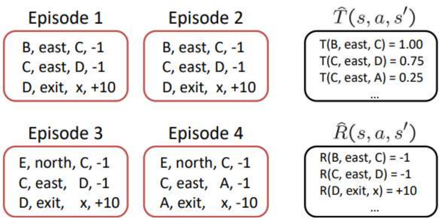
Policy Iteration : 정책은 V보다 빨리 수렴

- 1. 수렴할 때까지 Policy Evaluation
- 2. 수렴한 V값으로 Policy Extraction
- PI와 VI 모두 DP

RL : T나 R을 알지 못하는 상태의 Online Planning

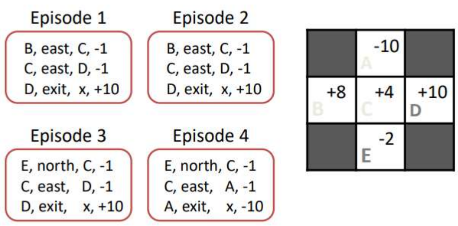
Model-Based : 에피소드에서 T와 R(모델)을 추측

- 에피소드의 각 스텝을 표집



Passive RL : 고정된 정책 $\pi(s)$를 통해 학습

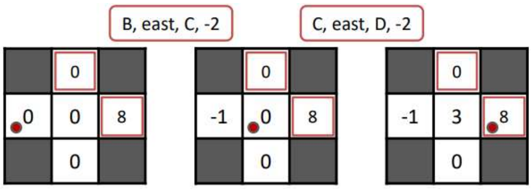
Direct Evaluation



- T와 R에 대한 지식이 필요 없음
- 상태 간 연결 정보를 무시함

TD Learning : 고정된 정책에서 샘플링

- $\mathrm{sample} = \mathrm{R}(s,\pi(s),s') + \gamma V^\pi(s')$
- $V^\pi(s) \leftarrow (1-\alpha)V^\pi(s) + \alpha \times \mathrm{sample}$



- 새로운 정책을 구하지 못함

Active RL : 최선의 정책을 구함

Q-Learning

- $\mathrm{sample} = \mathrm{R}(s,a,s') + \gamma \max_{a'} \mathrm{Q}(s',a')$
- $Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha \times \mathrm{sample}$
- Off Police Learning : 최적으로 행동하지 않아도 최적의 값을 학습
- Caveats : 충분한 탐색, 결과적으로 작아지는 LR

Exploration

- ϵ-Greedy : ϵ 확률로 무작위 행동
- Thrashing을 피하기 위해 ϵ를 점점 줄임
- Exploration Function : 나쁜 선택지를 선택 X
- $f(u,n) = u + \frac{k}{n}$
- $\mathrm{sample} = \mathrm{R}(s,a,s') + \gamma \max_{a'} f(\mathrm{Q}(s',a'),\mathrm{N}(s',a'))$

Approximate Q-Learning(Feature Based)

- $Q(s,a) = \sum wf(s,a)$
- $\mathrm{dif} = [\mathrm{r} + \gamma \max_{a'} \mathrm{Q}(s',a')] - \mathrm{Q}(s,a)$
- $Q(s,a) \leftarrow Q(s,a) + \alpha \times \mathrm{dif}$
- $w_i \leftarrow w_i + \alpha \times f_i(s,a) \times \mathrm{dif}$

랜덤변수 : Measurable Function $X: \Omega \rightarrow E$로 나타냄

- Measurable Space E : [0, 1], 합이 1

결합확률분포 : $P(x,y,z,\dots)$

주변확률분포 : $P(x,y,z,\dots)$를 $P(x)$로

조건부 확률 : $P(x|y) = \frac{P(x,y)}{P(y)}$

Product Rule : $P(x,y) = P(x|y)P(y)$

Chain Rule : $P(x_1,\dots,x_n) = \prod_i P(x_i|x_{<i})$

베이즈 정리 : $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(B,A)}$

$= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$

- 사전확률 : P(e)
- 조건부 확률, 우도 : P(D|e)
- 사후확률 : P(e|D)
- 사전확률로 사후확률을 구함

열거에 의한 확률추론 : $O(d^n)$의 시공간 복잡도

NB : 피쳐들이 레이블에 대해 조건부 독립

- MLE : $\theta = \arg\max_\theta P(X|\theta)$, 데이터 포인트의 우도 곱

- MAP : $\theta = \arg\max_\theta P(\theta|X)$, MLE로부터 값 유도
- 조건부 독립 : Z가 결정되면, Y가 X에 미치는 영향 X
- $P(x,y|z)=P(x|z)P(y|z)$ 또는 $P(x|y,z)=P(x,z)$
- $P(X,c_j)=P(X_1,...,X_k|c_j)P(c_j)=P(c_j)\prod_i^k P(X_i|c_j)$
- $P(c_j|X_1,...,X_k)$를 최대화하는 것이 목표
- $P(c_j)\prod_{i=1}^k P(X_i|c_j)$를 최대화하는 것과 같음
- $\log P(c_j)+\sum_{i=1}^k \log P(X_i|c_j)$를 최대화하는 것과 같음
- $P(c_j)=\frac{N_d(c_j)}{N_d}$(사전확률)
- $P(X_i|c_j)=\frac{N_w(X_i,c_j)}{N_w(c_j)}$(우도)
- Nd : 전체 문서의 수, Nd(c) : 클래스 c의 문서의 수
- Nw(X,c) : c의 문서에서 X가 등장하는 횟수
- Nw(c) : c의 문서에 있는 모든 피처의 수
- Laplace Smoothing : $p=\frac{N_1+\alpha}{N_2+\beta}$
- 장점 : 간단하고 빠름, 희소한 데이터에도 잘 작동
- 단점 : 오차, 편향의 영향, 데이터의 패턴을 못 찾음

결정트리 : 인간이 해석할 수 있는 모형
- 분기의 정지 : 모든 데이터가 같은 클래스, 남은 속성들 전부에 대해 데이터가 동등한 경우
- 엔트로피 : $-\sum P(Y)\log P(Y)$
- 조건부 엔트로피 : $-\sum P(X)\sum P(Y|X)\log P(Y|X)$

Bias-Variance Tradeoff
- 편향 : 훈련 오차 : 과소적합
- 분산 : 검정 오차 : 과적합
- 해결 : Validation Set 활용, 프루닝, 고정 높이, 고정 리프 노드

군집화

K-Means : K개의 군집 중심을 이동하는 군집화
- 데이터의 할당이 바뀌지 않을 때까지 반복
- 중심의 위치는 유한 시간 내에 수렴
- 각 반복에서 점 할당에 O(KN), 평균 구하기에 O(N)

거리함수 속성
- Symmetric : $\mathrm{dist}(x,y)=\mathrm{dist}(y,x)$
- Positivity : $\mathrm{dist}(x,y)\geq 0$
- Self Similarity : $\mathrm{dist}(x,y)=0 \Leftrightarrow x=y$
- Triangle Inequality: $\mathrm{dist}(x,y)+\mathrm{dist}(y,z)\geq \mathrm{dist}(x,z)$

목적함수 : $\min_\mu \min_C \sum_{i=1}^k \sum_{x\in C_i}|x-\mu_i|^2$
- C : $\min_c \sum_{j=1}^n |x_j-\mu_{c_j}|^2$
- Mu : $\min_\mu \sum_{i=1}^k \sum_{x\in C_i}|x-\mu_i|^2$
- x는 데이터, mu는 중심, C는 할당된 데이터
- $\mu_i=\frac{1}{|C_i|}\sum_{x\in C_i}x$에서 기울기가 0
- 단점 : 초기화 위치, 분류하지 못하는 경우 있음

Agglomerative
- Single Link : 가장 가까운 거리
- Complete Link : 가장 먼 거리
- Average Link : 가능한 모든 경우의 평균 거리

신경망
- 선형 함수가 비선형 활성함수로 연결된 형태
- $f=W_3\max(0,W_2\max(0,W_1x))$
- N-Layer : N개의 가중치를 가짐
- N-Hidden Layer : N개의 중간 레이어를 가짐
- 입력 레이어 : $x\in\mathbb{R}^k$ / 중간 레이어 : $h\in\mathbb{R}^l$
- 출력 : $y\in\mathbb{R}$ / 연결 : $W\in\mathbb{R}^{l\times k}$
- 파라미터 수 : $kl_1+l_1l_2+...$

활성함수
- 시그모이드 : $\frac{1}{1+e^{-x}}$, Gradient Vanishing
- 하이퍼볼릭 탄젠트 : $\tanh(x)$. Gradient Vanishing
- ReLU : $\max(0,x)$, Dead ReLU
- Leaky ReLU : $\max(0.1x,x)$, 0에서 미분불가능
- ELU : 0 이하에서 $\alpha(e^x-1)$
- Softmax : $\frac{\exp(x_i)}{\sum\exp(x_j)}$, Confidence Score(확률)

역전파 : 입력 x, y, 출력 z
- Local : 입력으로 들어오는 $\frac{\partial z}{\partial x},\frac{\partial z}{\partial y}$
- Upstream : 출력이 되돌아오는 $\frac{\partial L}{\partial z}$
- Downstream : 입력에 되돌아가는 $\frac{\partial L}{\partial x},\frac{\partial L}{\partial y}$

미분

분수함수 : $-\frac{1}{x^2}$

지수함수 : $a^x\ln a$, $e^x\rightarrow e^x$

로그함수 : $\frac{1}{x\ln a}$, $\ln x\rightarrow\frac{1}{x}$

패턴