https://github.com/multicore-it/n

강화학습 기본 알고려즘

2. 다이나믹 프로그래밍

https://github.com/multicore-lt/r

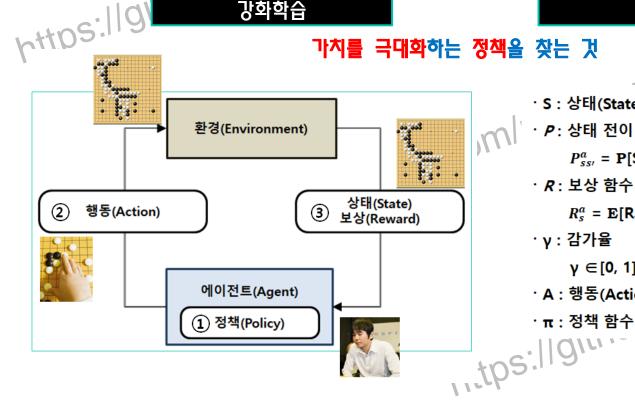
menticore-it/r/

기본개념



MDP





- · S : 상태(State)의 집합
- · P: 상태 전이 매트릭스

$$P_{ss'}^a = \mathbf{P}[\mathbf{S}_{t+1} = \mathbf{s}' \mid \mathbf{S}_t = \mathbf{s}, \mathbf{A}_t = \mathbf{a}]$$

· R: 보상 함수

$$R_s^a = \mathbf{E}[R_{t+1} \mid S_t = s, A_t = a]$$

· y : 감가율

 $\gamma \in [0, 1]$

- · A : 행동(Action)의 집합
- · m : 정책 함수



MDP 다시 보기^{ticore-itlrl} 가치계산 https://github.com

 $\mathbf{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ... \mid S_t = s]$

$$v_{\pi}(s) = \mathbf{E}_{\pi}[R_{t+1} + \gamma v]$$

$$v_{\pi}(s) = \mathbf{E}_{\pi}[R_{t+1} + \gamma v_{\pi}(S_{t+1}) \mid S_{t} = s]$$

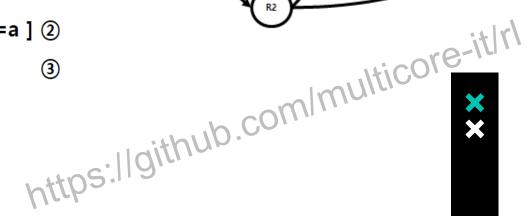
$$= \sum_{a \in A} \pi(a|s) R_{s}^{a} + \gamma \sum_{a \in A} \pi(a|s) \sum_{s' \in S} R_{s}^{a}$$

$$= \sum_{a \in A} \pi(a|s) R_s^a + \gamma \sum_{a \in A} \pi(a|s) \sum_{s' \in S} P_{ss'}^a v_{\pi}(s') \quad \text{1}$$

$$\text{1-2}$$

$$q_{\pi}(s, a) = \mathbf{E}_{\pi}[R_{t+1} + \gamma q_{\pi}(S_{t+1}, A_{t+1}) \mid S_{t}=s, A_{t}=a] ②$$

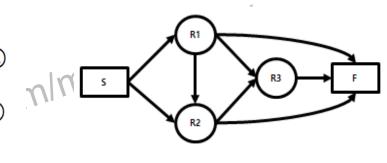
$$= R_{s}^{a} + \gamma \sum_{s' \in S} P_{ss'}^{a} \pi(s', a') q_{\pi}(s', a')$$
 ③



MDP 다시 보기^{ticore-itlrl} 정책결정 https://github.com

$$\pi^*(a|s) = \begin{cases} 1 & \text{if } a = \operatorname{argmax}_{a \in A} q * (s, a) \text{ } \\ 0 & \text{otherwise} \end{cases}$$

https://gru

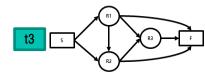


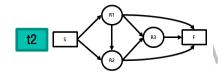
https://github.com/multicore-it/r/

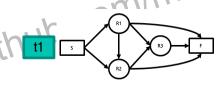
uticore-it/r/ 기본개념 itins://github.ct

기본개념

- 동적계획법, 환경에 대한 모든 정보를 알고 있는 모델기반(Model Based) 방법론 ithub.com/muli
- 모델기반 환경에서 사용하는 MDP 해결 방법
- 정책평가
 - 고정하고 처음 타임스텝과 뒤 따르는 스텝들에 대한 각각 구해서 합산
 - (2) 마지막 타임스텝까지 반복 수행
 - (3) 현재 타임스탭의 가치를 업데이트
- 정책제어
- 가치함수를 사용해서 탐욕적(greedy)으로 정책을 선택해서 현재 정책을 갱신(update)



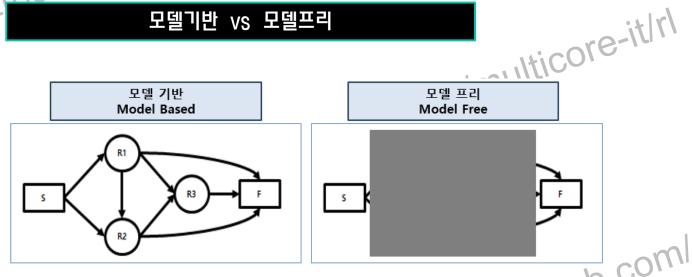






다이나믹 프로미백명 모델기반 vs 모델프리

모델기반 vs 모델프리



https://github.com/multicore-it/r/

마르코프 결정에바정 그리드 https://github.cata 그릭드월드

그리드월드

- 바둑판처럼 정사각형으로 나누어진 환경에서 에이전트가 목적지를 찾아가는
- 게임의 목적은 최단거리로 에이전트가 목적지를 찾아가도록 정책을 설정하는 것

0.22	0.25	0.27	0.30	0.34	0.38	0.34	0.30	0.34	0.38	
0.25	0.27	0.30	0.34	0.38	0.42	0.38	0.34	0.38	0.42	
0.22					0.46				0.46	4
0.20	0.22	0.25	-0.78		0.52	0.57	0.64	0.57	0.52	
0.22	0.25	0.27	0.25		0.08 R-1.	-0.36 R-1.0	0.71	0.64	0.57	
0.25	0.27	0.30	0.27		1.20 + R1.0	0.08 R-1.0	0.79	-0.29 R-1.0	0.52	
0.27	0.30	0.34	0.30		1.08	0.97	0.87	-0.21	0.57	
0.31	0.34	0.38	-0.58		-0.0B	-0.1β R-1.0	0.79	0.71	0.64	
0.34	0.38	0.42	0.46	0.52	0.57	0.64	0.7	0.64	0.57	
0.31	0.34	0.38	0.42	0.46	0.52	0.57	0.64	0.57	0.52	

					\cap	II	11.				
1	(50)	\Leftarrow	\Leftarrow	⇐	\Rightarrow	\Rightarrow	\Rightarrow	\Rightarrow	\Rightarrow	\Rightarrow	l
	₩			₩	Ĥ		Ü			∜	١
	↓		⇒				₩			↓	١
	↓		⇒	⇒	⇒				⇐	∜	١
	↓	⇒	⇒	⇒	⇒	⇒	⇒	∜	=	∜	١
	↓			\Rightarrow	\Rightarrow			⇒	₩		١
							₩	₩	₩		
	⇒	⇒	↓	⇐	¢=		₩	∜	₩		١
	\Rightarrow	\Rightarrow	(5*	⇐	=			∜	↓	₩	
		\Rightarrow	Î	⇐	=	⇐	⇐	=	¢		

Jub.com/multicore-it/r/

마르코프 결정에바정 그리드 https://github.cata

	1	k =	= 0	
(0.0	0.0	0.0	0.0
	0.0	목적	택지	0.0
	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0
			_5	라적지

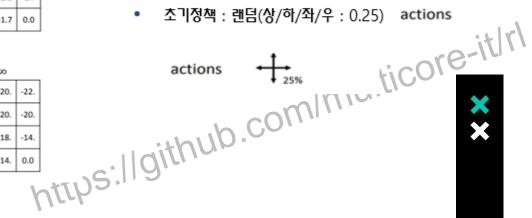
		_		
(2) k =	= 1	초기	화
0.0	-1.0	-1.0	-1.0	
-1.0	-1.0	-1.0	-1.0	
-1.0	-1.0	-1.0	-1.0	
-1.0	-1.0	-1.0	0.0	

(3) k =	2	좌표	Ē
0.0	-1.7	-2.0	-2.0	
-1.7	-2.0	-2.0	-2.0	
-2.0	-2.0	-2.0	-1.7	
-2.0	-2.0	-1.7	0.0	

multicore-it/r ±(0,1)

- 상태전이확률 : 1로 가정
- 보상: 타임스텝에 따라 -1
- 초기정책: 랜덤(상/하/좌/우: 0.25) actions

$k = \infty$						
0.0	-14.	-20.	-22.			
-14.	-18.	-20.	-20.			
-20.	-20.	-18.	-14.			
-22.	-20.	-14.	0.0			



기르코프 결정이바정 그리드 https://github.catagore-it/n/

상태가치계산

nulticore-it/r/

정책평가

http-.l/github.com/multicore-it/r/ -1.0 + (0.0*0.25 + -1.0*0.25 + -1.0*0.25 + -1.0*0.25) = -1.75

현재상태의 가치

1

2

3

마르코프 결정에바정 그리드 그리드월드

정책 업데이트

정책평가			정책	제어	
0.0 -1420	-22.	1	2	3	4
-141820	-20.	5 †	<u> </u>	7	8 †•
-202018	-14.	9 †	10	11 • †	¹² , •
-222014	0.0	13 t	¹⁴ →	¹⁵ →	16

-20.

-20.

-18.

-14.

-20.

-20.

가치가 가장 큰 그리드로 이동하도록 정책 설정

정책제어

om/multicore-it/r/

$$\pi^*(a|s) = \begin{cases} 1 & \text{if } a = \operatorname{argmax}_{a \in A} q * (s, a) \text{ } \text{ } \text{ } \\ 0 & \text{otherwise} \end{cases}$$

nttps://githur

정책평가 정책제어

마르코프 결정이 바정 그리드. https://github.care-it/r/ 그리드월드

https://cs.stanford.edu/people/karpathy/remforcejs/gridworld_dp.html

https://github.com/multicore-it/r/