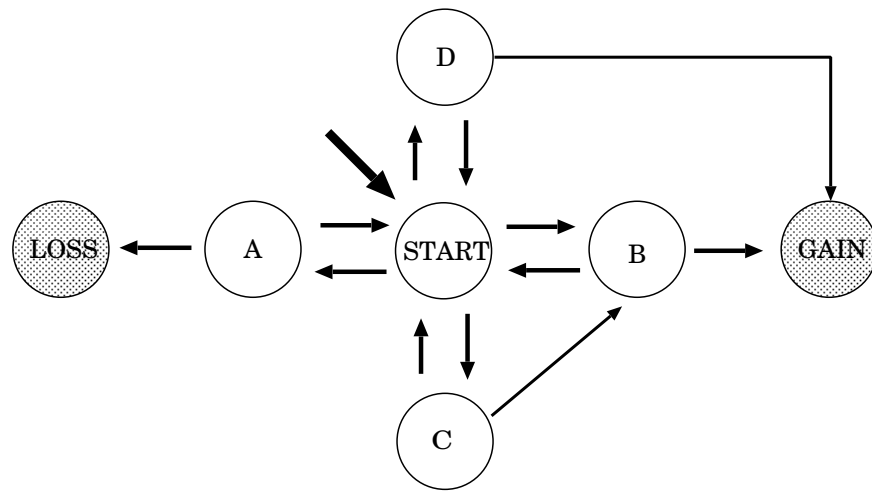


人工知能統論課題 10

[課題 10]

[問 10-1] (60 点)

以下の図に示す MDP 問題を考える．この問題では，LOSS 状態と GAIN 状態を吸収状態とし，GAIN に到着した場合に報酬 50，LOSS に到着した場合に報酬 -10 を与え，他の状態では報酬はない．各状態での可能な遷移は矢印で示す．本問題に対し，状態 START から開始し，割引率 (減衰率) 0.4 の Q-Learning を適用した場合，収束状態での Q テーブルを作成せよ．



State	Action	Q-value
START	←	
START	↑	
START	→	
START	↓	
A	←	
A	→	
B	←	
B	→	
C	↑	
C	↗	
D	↓	
D	↘	

[問 10-2](10 点)

Q-学習が収束したときに得られる制御ポリシーを示せ．