6509611684 ณัชชา สุขะเมือง

Homework: การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning)

ต้องการหยิบไพ่สองหรือสามใบจากไพ่สำรับหนึ่ง ให้เลขในหลักหน่วยมีค่าใกล้เคียงเลข 9 มากที่สุด
จงเขียนสถานะ (State) และการเปลี่ยนสถานะ (Transition) ของการหยิบไพ่ โดยให้เครื่องเรียนรู้ว่าควรมีการหยิบไพ่ใบที่สามหรือไม่ เพื่อให้จำนวนสถานะ (State) ลดลงและง่ายต่อการทำข้อสอบ กำหนดให้เงื่อนไขที่เครื่องจะใช้การเรียนรู้แบบเสริมกำลังในการตัดสินใจ คือ ไพ่สองใบแรกที่หยิบได้มีเลขรวมกันในหลักหน่วย เท่ากับ 5 เท่านั้น

- ถ้าไพ่สองใบแรกที่หยิบได้ มีเลขรวมกันในหลักหน่วย เท่ากับ 0, 1, 2, 3 และ 4 ให้เลือกหยิบไพ่ใบที่สาม
- ถ้าไพ่สองใบแรกที่หยิบได้มีเลขรวมกันในหลักหน่วย เท่ากับ 6, 7, 8 และ 9 ให้เลือกไม่หยิบไพ่ใบที่สาม

กำหนดให้ Unit Value เริ่มต้นของทุกสถานะ (State) เท่ากับ 1

ค่า Reward R_{t-1} เมื่อชนะ = 0.10 และ เมื่อแพ้ = -0.10

ค่า Reward R_{t-2} เมื่อชนะ = 0.05 และ เมื่อแพ้ = -0.05

ฟังก์ชันการเรียนรู้ของ Unit Value:

$$U_t = (U_{t-1} + R_{t-1}) \quad \text{เมื่อ } t<=2$$

$$= \frac{(U_{t-1} + R_{t-1}) + (U_{t-2} + R_{t-2})}{2} \quad \text{เมื่อ } t>2$$

โดยให้ t แทน ครั้งที่ 1, 2, 3, …

ผลลัพธ์จากการเรียนรู้ก่อนหน้า เมื่อหยิบไพ่สองใบแรกมีเลขรวมกันในหลักหน่วยได้เท่ากับ 5 เป็นดังนี้

1) เลือกไม่หยิบไพ่ใบที่สาม แล้ว แพ้
2) เลือกหยิบไพ่ใบที่สาม ได้เลข 3 แล้ว ชนะ
3) เลือกหยิบไพ่ใบที่สาม ได้เลข 6 แล้ว แพ้
4) เลือกหยิบไพ่ใบที่สาม ได้เลข 2 แล้ว แพ้
5) เลือกหยิบไพ่ใบที่สาม ได้เลข 1 แล้ว แพ้
6) เลือกไม่หยิบไพ่ใบที่สาม แล้ว ชนะ
7) เลือกไม่หยิบไพ่ใบที่สาม แล้ว ชนะ

ในครั้งถัดไป สมมติเครื่องได้ไพ่สองใบแรกมีเลขรวมกันในหลักหน่วยเท่ากับ 5 เครื่องจักรจะเลือกหยิบหรือไม่หยิบไพ่ ใบที่สาม? จะต้องเลือกไม่หยิบไพ่แล้วแพ้อีกกี่ครั้ง เครื่องจึงจะเปลี่ยนการตัดสินใจไปเลือกหยิบไพ่ใบที่สามแทน สมมติให้การตัดสินใจของเครื่องใช้วิธีแบบ Greedy

	Unit Value ของสถานะหยิบ	Unit Value ของสถานะไม่หยิบ
t=0	1.000	1.000
t=1	1.000	0.900
t=2	1.100	0.900
t=3	0.9750	0.900
t=4	0.9625	0.900
t=5	0.8938	0.900
t=6	0.8938	0.9750
t=7	0.8938	1.0125
t=8	0.8938	0.9188

ตอบ - ในครั้งที่ 8 ถ้าเครื่องได้ไพ่สองใบแรกมีเลขรวมกันในหลักหน่วยเท่ากับ 5 เครื่องจักรจะเลือก “ไม่หยิบไพ่ใบที่สาม” เพราะว่า เพราะ Unit Value ของสถานะไม่หยิบ (0.9188) สูงกว่า Unit Value ของสถานะหยิบ (0.8938)

- ต้อง แพ้จากการไม่หยิบ 3 ครั้งติดกัน (ติดต่อกันในรอบ t=8,9,10)
เครื่องจึงจะเปลี่ยนใจไปเลือก “หยิบไพ่ใบที่สาม” ทันทีในรอบถัดไป ตามวิธีแบบ Greedy