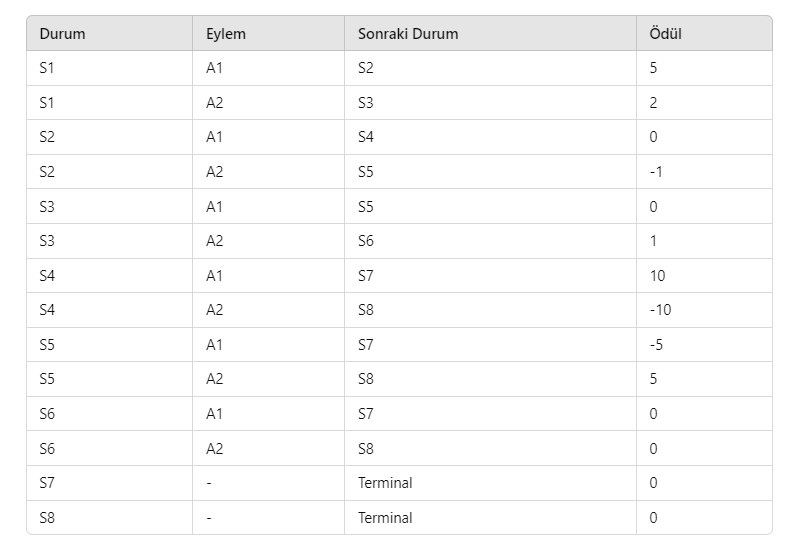
|  |
| --- |
| S O R U L A R |

1. Bir robot 8 durumlu bir yüzeyde hareket etmek istemektedir.

* Durumlar (S): S1, S2, S3, S4, S5, S6, S7, S8
* Eylemler (A): A1, A2 (Her durumda iki eylem seçeneği var.)
* İndirim Faktörü (&): 1

“Geçişler ve Ödüller” tablosu aşağıda verilmektedir. “Terminal” durumuna ulaştığında robot daha fazla eylem gerçekleştiremez ve hareket sonlanır, robot artık başka bir eylem yapamaz. Başlangıç durumu S1’dir ve geçiş olasılıkları görmezden gelinmiştir.  
  


1. Q değerleri için **denklem sistemini yazınız** ? (20P)
2. V değerleri için **denklem sistemini yazınız** ? (20P)
3. Başlangıç durumundan başlayarak (S1) son duruma kadar (Terminal), kumülatif toplam ödülü maksimize eden **yolu gösteriniz**, ödül değerini **hesaplayınız** ? (20P)
4. Q(S1,A1)=R(S1,A1)+&max(S1,A1’)

Q(S1,A2)=R(S1,A2)+&max(S1,A2’)

Q(S2,A1)=R(S2,A1)+&max(S2,A2’)

Q(S2,A2)=R(S2,A2)+&max(S2,A2’)

Q(S3,A1)=R(S3,A1)+&max(S3,A1’)

Q(S3,A3)=R(S3,A2)+&max(S3,A2’)

Q(S4,A1)=R(S4,A1)+&max(S4,A1’)

Q(S4,A2)=R(S4,A2)+&max(S4,A2’)

Q(S5,A1)=R(S5,A1)+&max(S5,A1’)

Q(S5,A2)=R(S5,A2)+&max(S5,A2’)

Q(S6,A1)=R(S6,A1)+&max(S6,A1’)

Q(S6,A2)=R(S6,A2)+&max(S6,A2’)

Q(S7,A1)=R(S7,A1’)=0

Q(S7,A2)=R(S7,A2’)=0

Q(S8,A1)=R(S8,A1’)=0

Q(S7,A2)=R(S7,A2’)=0

1. V(S1)= max(R(S1,A1)+&V(S2),/+ R(S1,A2)+&V(S3)}

V(S2)= max(R(S2,A1)+&V(S4),/+ R(S2,A2)+&V(S5)}

V(S3)= max(R(S3,A1)+&V(S5),/+ R(S3,A2)+&V(S6)}

V(S4)= max(R(S4,A1)+&V(S7),/+ R(S4,A2)+&V(S8)}

V(S5)= max(R(S5,A1)+&V(S7),/+ R(S5,A2)+&V(S8)}

V(S6)= max(R(S6,A1)+&V(S7),/+ R(S6,A2)+&V(S8)}

V(S7)=R(S7,A1)

V(S8)=R(S8,A2)

1. S1 (A1)🡪S2(A1)🡪S4(A1)-->S7, 5+0+10=15

Adım 1:

* Mevcut Durum: S13
* Eylem: Yukarı (çünkü yukarı hareket edebilir)
* Yeni Durum: S9
* Alınan Ödül: -1 (hareket maliyeti)
* Toplam Ödül: -1

Adım 2:

* Mevcut Durum: S9
* Eylem: Yukarı
* Yeni Durum: S5
* Alınan Ödül: -1
* Toplam Ödül: -1 + (-1) = -2

Adım 3:

* Mevcut Durum: S5
* Eylem: Yukarı
* Yeni Durum: S1
* Alınan Ödül: -1
* Toplam Ödül: -2 + (-1) = -3

Adım 4:

* Mevcut Durum: S1
* Eylem: Yukarı (artık yukarı hareket edemez, çünkü üst sınırda)
* Alternatif Eylem: Sağa
* Yeni Durum: S2
* Alınan Ödül: -1
* Toplam Ödül: -3 + (-1) = -4

Adım 5:

* Mevcut Durum: S2
* Eylem: Yukarı (yine yukarı hareket edemez)
* Alternatif Eylem: Sağa
* Yeni Durum: S3
* Alınan Ödül: -1
* Toplam Ödül: -4 + (-1) = -5

Adım 6:

* Mevcut Durum: S3
* Eylem: Yukarı (yukarı hareket edemez)
* Alternatif Eylem: Sağa
* Yeni Durum: S4 (hedef durumu)
* Alınan Ödül: -1 (hareket maliyeti) + +10 (hedefe ulaşma ödülü)
* Toplam Ödül: -5 + (-1) + 10 = +4

Sonuç:

Robot, belirlenen deterministik politikayı izleyerek S13'ten S4'e toplam 6 adımda ulaşmıştır. Her adımda aldığı hareket maliyeti nedeniyle toplamda -6 ödül almış, ancak hedefe ulaştığı için +10 ödül kazanmıştır.

* Toplam Ödül:
  + Hareket Maliyetleri: -6
  + Hedef Ödülü: +10
  + Net Toplam Ödül: +4

1. Proje
2. Önce exploration, sonra exploitation, ve ilerleyen süreçte bunlar arasında bir denge oluşturmaya çalışır.