

## Q-learning ile Yol Planlaması

---

Proje İlan Tarihi  
Proje Teslim Tarihi:

21 Nisan 2021  
17 Mayıs 2021

### 1-) Giriş

Pekiştirmeli öğrenme (reinforcement learning), öznelerin (agent) bir görevi en yüksek kazançla tamamlayabilmek için hangi eylemleri gerçekleştirmeleri gerektiği ile ilgilenen bir makine öğrenmesi tekniğidir. Bu tür öğrenme algoritmalarının girdisi öznelerin görev yapacakları farklı durumlardan oluşan bir ortam  $S$ , yapabilecekleri eylemler  $A$ , ortamdaki durumuna göre yapabilecekleri eylemleri belirleyen prensipler, bir durumdan diğer duruma geçtiklerinde elde edecekleri kazançtır.

Q-learning pekiştirmeli bir öğrenme algoritmasıdır. Ortam hakkında hiçbir şeyin bilinmediği durumlarda, Q-learning algoritması ortamı brute-force şeklinde, her ortam için olası tüm aksiyonları takip ederek, problem çözümü için en karlı yolu bulmaya çalışır. Q-learning algoritmasının girdileri kazanç matrisi olarak adlandırılan  $R$  matrisidir. Bu matrisin satır ve sütunları ortamları temsil etmekte,  $R[i][j]$  değeri ise  $i$  durumundan  $j$  durumuna geçtiğinde elde edilen anlık kazanç değeridir. Eğer  $i$  durumunda  $j$  durumuna bir geçiş yoksa  $R[i][j]$  değeri -1, geçiş var ancak  $j$  durumu hedef durum değilse değeri 0,  $j$  hedef durum ise değeri kullanıcı tarafından belirlenen bir kazanç değeridir.

Q-learning algoritmasının çıktısı ise öğrenmenin kalitesini gösteren  $Q$  matrisidir. Q-learning iteratif bir algoritmadır ve tüm değerleri başlangıçta 0 olan  $Q$  matrisi optimal değerlere yakınsadığı da sona erer. Algoritma her iterasyonda rastgele bir durumdan öğrenmeye baslar,  $A$ 'ya göre durum değiştirir ve  $Q$  matrisini günceller.  $A$ 'ya göre hedef duruma ulaşıldığında iterasyon sona erer.  $A$ 'ya göre bir durumdan birden fazla duruma geçiş olabilir. Böyle bir durumda, olası geçişler den biri rastgele seçilir. Eğer seçilen durum hedef duruma ulaştırmıyorsa, durum rastgele olacak durum olarak belirlenir. Hedef duruma ulaşılan

$$R = \begin{pmatrix} -1 & 0 & -1 & 0 & -1 & -1 \\ 0 & -1 & 100 & -1 & 0 & -1 \\ -1 & -1 & 0 & -1 & -1 & -1 \\ 0 & -1 & -1 & -1 & 0 & -1 \\ -1 & 0 & -1 & 0 & -1 & 0 \\ -1 & -1 & 100 & -1 & 0 & -1 \end{pmatrix}$$

Başlangıç durumu  $A$  olarak seçilsin ve  $\gamma = 0.8$  olsun.  $A$  durumundan olası aksiyonlar  $B$  durumuna geçiş ve  $D$  durumuna geçişidir.. Rastgele olarak  $B$  durumuna geçiş seçilsin. Buna göre

$$Q(0, 1) = R(0, 1) + 0.8 \times \text{Max}\{Q(1, 4), Q(1, 2)\}$$

Buna göre  $Q(0,1) = 0$  olarak güncellenir.  $B$  durumu hedef durum olmadığı için, durum  $B$  olarak güncellenir ve iterasyon devam eder.  $B$  durumundan olası aksiyonlar  $C$  durumuna geçiş ve  $E$  durumuna geçişidir. Rastgele olarak  $C$  durumuna geçiş seçilsin. Buna göre

$$\underline{Q(1, 2) = R(1, 2) + 0.8 \times \text{Max}\{Q(2, 2)\}}$$

Bu formülünün farklı türevleri olmakla birlikte ödevde bu formülü kullanınız.

Buna göre  $Q(1,2) = 100$  olarak guncellenir ve  $C$  hedef durum olduğu için iterasyon sonlandırılır. Optimuma yakınsama oluşana kadar yeni iterasyonlar işletilir.

## 2-) Ödev

Burada robotun Q learning algoritması kullanarak engel sütunlarından kaçması ve beyaz alanlardan geçerek doğru yol alması gerekiyor. Aşağıdaki verilen matrisleri gerçek ortamdaki bir yol olarak düşünün. Robotumuz mavi kareden başlayıp kırmızı kutulara çarpmadan bitiş kısmına en kısa(maliyetle) yoldan ulaşırsa başarılı sayılacaktır.

Ajan, herhangi bir beyaz kareden başlayarak sağa, sola, aşağı, yukarı ve çapraz hareket edebilir. Atılan adımlar belirleyici olmalı ve engele çarpışmadıkça başarılı olur. Robot en son duvara geldiğinde robot sadece aşağı hareket ederek istenilen noktaya “37” gelecektir. Sonuç olarak robot başlangıç noktasından istenilen hedefe gelinceye kadar hiçbir engele çarpmadan ve en kısa yolu bularak ödülü alır. Rr: ajan[1,2, 8] karelere çarparsa işlem bitirir. Aksi takdirde, diğer her kareden herhangi bir işlem yapmak rs ödüllendirilir.

	1	9	17	25	33
	2	10	18	26	34
	3	11	19	27	35
	4	12	20	28	36
	5	13	21	29	37
	6	14	22	30	38
	7	15	23	31	39
	8	16	24	32	40

Örnek matris tablosu.

İndirim faktörü  $\gamma = 0.9$ , kırmızıya çarparsa -5 ödül puanı, yeşil bitiş noktasına +5, diğer geçişlere +3 ödül puanı olarak hesaplanacaktır.

- Verilen 50 \* 50'lik matriste her bir kullanıcı kendine özgü engel oluşturup, matristeki değerleri random olarak atayacaktır. Bu matris değerlerini engel.txt dosyasına yazdırılacak. Örnek gösterim (1,1,K)(10p).
- Grafiksel ara yüzde belirlenen yollar, engeller ve duvarlar gösterilecektir(15p).
- Kullanıcı tarafından bir grafiksel arayüz tasarlanacak, bu ara yüzde ajan başlangıç noktası, hedef noktası istenecektir. (15p)
- Herhangi bir başlangıç noktasından hedef noktaya ulaşmaya kadar ajanın yaptığı kazançların/maliyetin(episode via cost) ve bölüm adım sayısının (episode via step) grafiği çizdirecek (30p).
- Sonuç olarak ise başlangıç karesinden hedef kareye giden en kısa yol grafiksel ara yüzde gösterilerek yol planı grafik üzerinde çizdirilecek.. (30p).

### Ödev Teslimi

- Dersin takibi projenin teslimi dahil edestek.kocaeli.edu.tr sistemi üzerinden yapılacaktır. edestek.kocaeli.edu.tr sitesinde belirtilen tarihten sonra getirilen projeler kabul edilmeyecektir.
- Proje ile ilgili sorular edestek.kocaeli.edu.tr sitesindeki forum üzerinden Arş. Gör. Yılmaz DİKİLİTAŞ'a veya Arş. Gör. M.Ahmet DEMİRTAŞ'a sorulabilir.
- Sunum sırasında algoritma, geliştirdiğiniz kodun çeşitli kısımlarının ne amaçla yazıldığı ve geliştirme ortamı hakkında sorular sorulabilir. Kullandığınız herhangi bir satır kodu açıklamanız istenebilir.
- Sunum tarihleri daha sonra ayrıca duyurulacaktır.

- Proje grupları 2 kiřiden oluřmalıdır. Proje grup bilgileri ařaęıda verilen linke en ge 17 Mayıs Pazartesi gnne kadar girilmelidir. Bu tarihten sonra gruplarda herhangi bir deęiřiklik yapılmayacaktır. Link:

[https://docs.google.com/spreadsheets/d/105UOdbacsZf\\_vKNO-o4wxNTMRyTtSE2EYrmqsspjWSQ/edit?usp=sharing](https://docs.google.com/spreadsheets/d/105UOdbacsZf_vKNO-o4wxNTMRyTtSE2EYrmqsspjWSQ/edit?usp=sharing)

- Projenin tanıtım toplantısı 23.04.2021- Cuma gn saat 15:00'te blm duyurularında ve e-destekte duyurulacak toplantı linki zerinden uzaktan yapılacaktır.