ELM472 Makine Öğrenmesinin Temelleri Pekiştirmeli Öğrenme Ödev-4

Selimhan Aygün
s.aygun2019@gtu.edu.tr
Elektronik Mühendisliği Bölümü, GTÜ, Kocaeli, Türkiye

ÖZET

Bu çalışmada pekiştirmeli öğrenme q-learning metodu kullanılarak uygulanmıştır.

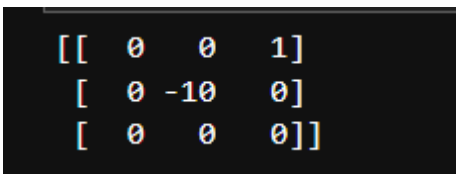
GİRİŞ

Pekiştirmeli öğrenme, bir makinenin çevresiyle etkileşim kurarak ve elde ettiği deneyimleri kullanarak kendini geliştirdiği bir makine öğrenmesi yaklaşımıdır. Pekiştirmeli öğrenme, bir ajanın, bir ortamda belirli bir hedefi elde etmek için en uygun eylemleri seçmeyi öğrendiği bir öğrenme paradigmasıdır. Bu çalışmada kullanılan Q-learning, pekiştirmeli öğrenmenin en çok bilinen algoritmalarından biridir. Q-learning algoritması, bir sonraki hareketleri inceleyip yapacağı hareketlere göre kazanacağı ödülü görmek ve bu maksimize etmeyi ve buna göre hareket etmeyi amaçlar.

TEORİ VE YÖNTEM

Verilen probleme göre çevrenin tanımlaması yapıldı. Çevre 3 ana bileşenden oluşmaktadır. Durum (state), eylem (action) ve ödül (reward). Durumlar ve eylemler Q-öğrenme yapay zeka aracısı için girdilerdir; olası eylemler ise yapay zeka aracısının çıktılarıdır.

Durumlar, çevrede bulunan tüm olası yerlerdir. Başlangıç noktası (B), beyaz kutular, bitiş noktası (T), gri olarak verilen kutunun tamamı birer durumdur. B ve T birer terminal durumudur.
Eylemler, ajanın haraket edebileceği yönlerdir. Bu durumda “up”, “down”, “right”,”left” olmak üzere 4 adet tanımlanmıştır.
Ödüller ise ajanın haraketleri sonucu aldığı geri dönütlerdir. Beyaz kutular 0, Bitiş noktası 1 ve gri kutu -10 olarak tanımlanmıştır.

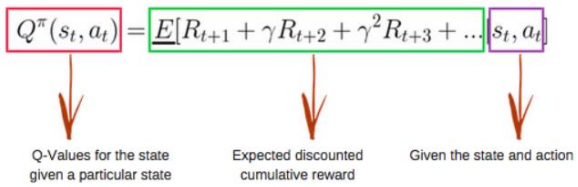


Şekil 1. Çevrenin görünümü

Q-Tablosu, her durumdaki eylem için gelecekte beklenen maksimum ödülleri hesapladığımız basit bir arama tablosudur. Temel olarak bu tablo bizi her durumdaki en iyi eyleme yönlendirecektir.
Q-tablosunun satırları durumlar sütunları ise eylemlerdir. Her Q-tablosu puanı, ajanın bu durumda o eylemi gerçekleştirmesi durumunda alacağı gelecekte beklenen maksimum ödül olacaktır. Bu yinelemeli bir süreçtir, çünkü her yinelemede Q-Table'ı geliştirmemiz gerekir. Tablo 1’ de problemimiz için örnek tablo görülebilir.

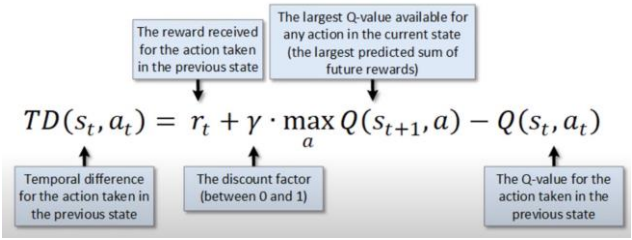
Tablo 1. Q-tablosu

Eylemler	UP	RIGHT	DOWN	LEFT
Başlangıç				
Beyaz				
Gri				
Bitiş				



Şekil 2. Q- fonksiyonu

Yukarıdaki fonksiyonu kullanarak tablodaki hücreler için Q değerlerini elde edilir. Başlangıçta Q tablosundaki tüm değerler sıfırdır. Değerlerin güncellenmesi yinelenen bir süreç vardır. Çevreyi keşif başladığında Q fonksiyonu, tablodaki Q değerlerini sürekli olarak güncelleyerek daha iyi ve daha iyi yaklaşımlar sağlar. Yeteri kadar denemeden sonra bilgiden istifade(exploitation) etmeye başlar. Bu aşamayı Epsilon-Greedy algoritması kullanarak gerçekleştirmektedir.



Şekil 3. TD fonksiyonu

Şekil 3'teki TD fonksiyonu formülü ajanın belirli bir durum ve eylem için beklenen ödülünü hesaplamak için kullanılır. TD fonksiyonu, mevcut durumda alınan ödülü ve gelecekteki maksimum Q-değerini dikkate alarak bu beklentiyi günceller.

```
Training complete!
Shortest Path: [[2, 0], [1, 0], [0, 0], [0, 1], [0, 2]]
```

Şekil 4. Sonuç

III. ANALİZ VE YORUM

Bu çalışmada pekiştirmeli öğrenme ve q-learning metodu başarıyla gerçekleştirilmiştir. Ancak, en kısa yolun bir alternatifi olmasına rağmen hep aynı sonuç elde edilmiştir. Random'dan kaynaklanıldığı düşünülse de see değiştirilmesine rağmen aynı sonuç elde edildi. Bir çözümü bulunamadı.

IV. KAYNAKÇA

[1] https://www.freecodecamp.org/news/an-introduction-to-q-learning-reinforcement-learning-14ac0b4493cc/

[2]https://colab.research.google.com/drive/1E2RViy7xmor0mhqskZV14_NUj2jMpJz3