6×6 Taxi Ortamında Optimize Q-Learning

Elif Yılmaz - 25435004004

Özet

Bu çalışmada klasik Taxi-v3 ortamı 6×6 boyutunda engeller içerecek şekilde yeniden tasarlanmış, action masking uygulanmış ve durum uzayı gereksiz bilgilerin kaldırılmasıyla 622.080'den **47.952** duruma indirgenmiştir. 5 milyon episodluk Q-Learning eğitimi sonucunda ajan ortalama 12 adımda görevi tamamlamayı öğrenmiş ve en iyi ortalama ödül **8.99** olarak elde edilmiştir.

1 Giriş

Taxi-v3 ortamı küçük yapısı nedeniyle tabular yöntemlerle kolay çözülebilmektedir. Bu çalışmada daha zor bir 6×6 grid-world tasarlanmış, engeller eklenmiş ve dinamik başlangıç koşullarıyla problem karmaşıklaştırılmıştır. Amaç, action masking ve durum uzayı sadeleştirmesinin tabular Q-Learning performansına etkisini incelemektir.

2 Ortam Tasarımı

- Harita: 6×6, 3 yasak hücre ve 4 sabit duvar.
- Yolcu ve hedef her reset'te rastgele belirlenir.
- Pickup/dropoff tüm geçerli hücrelerde yapılabilir.
- Action masking ile duvara çarpma ve geçersiz işlemler engellenir.

Aksiyonlar: yukarı, aşağı, sol, sağ, pickup, dropoff (6 adet)
Ödüller: normal adım –1, geçersiz hareket –10, başarılı bırakma +20.

Durum Uzayı

Yolcu taksideyken konumu ayrıca tutulmadığı için:

- Yolcu dışarıda: $6^4 = 46{,}656$
- Yolcu takside: $6^2 \times 6^2 = 1{,}296$

$$S_{\text{total}} = \mathbf{47{,}952}$$

Bu sadeleştirme durum uzayını yaklaşık 13 kat küçültmüştür.

3 Yöntem

- Algoritma: Q-Learning
- Hiperparametreler: $\alpha = 0.1$, $\gamma = 1.0$, $\epsilon = 0.1$
- Eğitim: 5.000.000 episod, her 50.000'de değerlendirme
- Action masking tüm aşamalarda aktif
- En iyi politika `best_q_table.npy` dosyasına kaydedildi

4 Eğitim Sonuçları

Episod	Ortalama Ödül	Adım
50.000	–33.26	54.3
100.000	+3.85	17.1
500.000	+8.94	12.1
2.350.000	**+8.99**	12.0
5.000.000	+8.98	12.0

Tablo 1: Eğitimdeki önemli kilometre taşları

Ajan yaklaşık 150 bin episotta görevi yapabilir hâle gelmiş, 2.3–2.6 milyon aralığında en iyi performansa ulaşmıştır. Eğitim süresi toplam **1 saat 8 dakika 49 saniyedir**.

5 Test Sonuçları

Senaryo	Adım	Ödül
1	25	+8
2	19	+12
3	25	+8

Tablo 2: Rastgele test görevlerinin sonuçları

Ajan tüm testlerde engellere takılmadan, yasak hücrelere girmeden ve optimuma yakın rotalar kullanarak görevi başarıyla tamamlamıştır.

6 Sonuç

Bu çalışmada action masking kullanıldı ve yolcu taksiye bindiğinde yolcu konumu durumdan çıkarılarak durum uzayı yaklaşık 13 kat küçültüldü. Bu iki basit değişiklik sayesinde 6×6 engelli Taxi ortamı 5 milyon episodda eğitilerek ortalama 12 adımda çözülebilir hâle geldi (en iyi ortalama ödül 8.99). Elde edilen sonuçlar, küçük ama etkili ön işlemlerle tabular Q-Learning'in hâlâ oldukça kullanışlı olduğunu göstermektedir.