Reinforcement Learning Türkiye 1. Ödev

Muhammet KARA (discord : kara)

Uğur İPEKDÜZEN (discord : u.ipkdzn)

Algoritma:

1. Kısım

173-177: Modeli eğitmek için kullanacak iki adet network oluşturulur ve daha sonra bu network'leri güncellemek için parametreleri değişkenlere atanır. (Q-network ve Q-target)

180-184: Bir sonraki state-action pair'den gelecek en yüksek Q-value hesaplanır.

186-187: Target value ve current estimate hesaplanır.

188: En son Bellman error hesaplanır.

2. Kısım

258: Şu anki frame'i replay buffer'a kaydedilir.

260-263: Rastgele sayı üretip bu sayı epsilon değeriyle karşılaştırılır. Eğer epsilondan küçükse veya model initialize edilmemişse veya replay bufferda sample edilecek kadar yeterince veri birikmemişse rastgele bir action seçilir.

265-267: Eğer rastgele değer epsilondan büyükse, replay bufferdan bir state alıp onu Q-network'ten geçirerek elde edilen sonuçlara göre en iyi action seçilir.

269: Environment'ta seçilen action uygulanır ve yeni elde edilen değerler değişkenler güncellenir.

270: Yeni değerler replay buffer'a eklenir.

272-275: En son state oyunun sonuysa environment resetlenir, değilse son state güncellenir.

3. Kısım

##3.a

323: Replay buffer'dan bir sample alınır.

##3.b

327: Model initialize edilmemişse initialize edilir ve "self.model_initialize" değişkeni True olarak güncellenir.

335: Learning rate parametresi ilgili değişkene atanır.

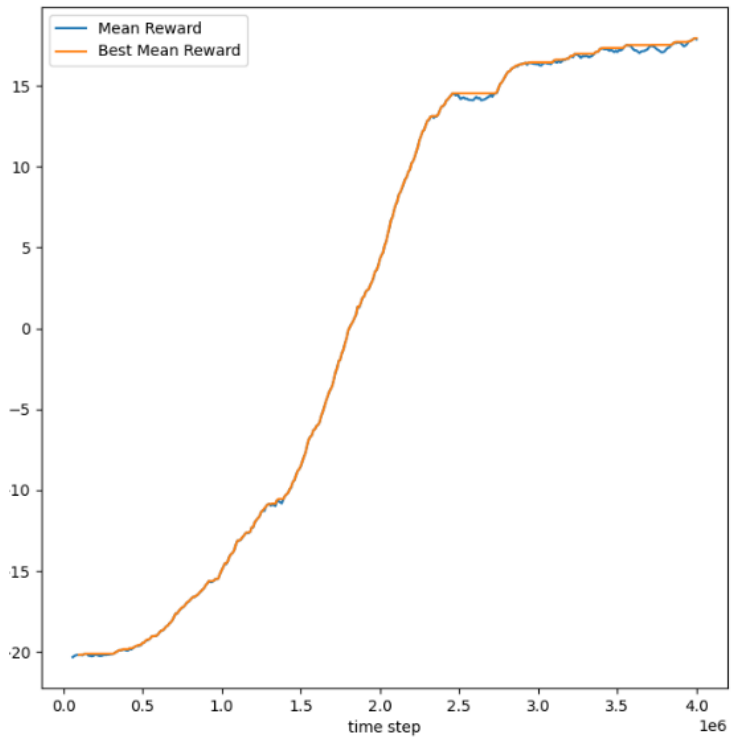
##3.c

338: Session'ı çalışırtırarak Q-network eğitilir.

#3.d

348-349: Eğer Q-target'in güncellenme frekansı geldiyse, Q-target güncellenir.

Soru 1)



Soru2) İncelenecek hyperparameter olarak learning rate seçilmiştir ve 3 farklı değer ile tekrar eğitilmiştir.

Sonuç olarak varsayılan değer olan 1.0e-4 yerine, 1.15e-4 olarak değiştirilmiş değer hedeflenen fonksiyona daha hızlı yaklaşmıştır.