Michael Jonathan Halim - 13521124 - GAIB Medium - Q-Learning

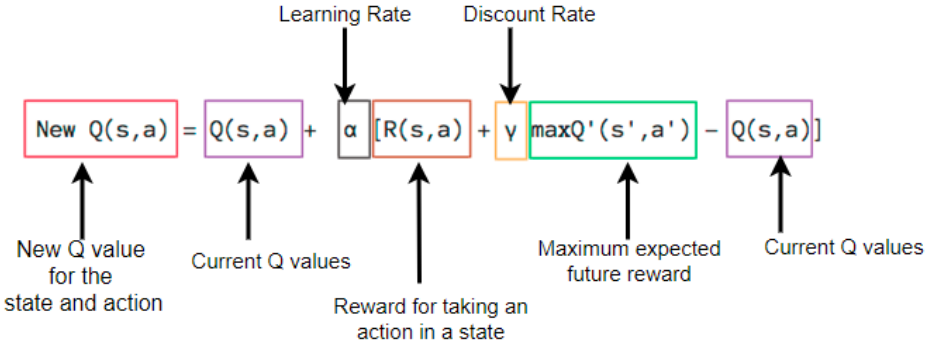
1. Jelaskan apa yang dimaksud dengan Reinforcement Learning!

   Reinforcement learning adalah salah satu jenis machine learning dimana terdapat suatu agen yang belajar untuk memilih keputusan terbaik dengan berinteraksi dengan lingkungan yang disediakan. Tujuan dari agen adalah mencapai target yang ditentukan dan mendapatkan feedback berdasarkan reward atau penalti dari lingkungan pembelajarannya. Maka dari itu, agen akan terus belajar mencapai target berulang kali untuk mendapatkan reward yang lebih baik dari sebelumnya.

2. Jelaskan bagaimana proses dari Q-learning bekerja!

   Q-Learning merupakan salah satu reinforcement learning tanpa mengetahui probabilitas transisi dan reward yang tepat. Oleh karena itu, Q-Learning memiliki tujuan untuk menemukan pergerakan yang optimal dengan memaksimalkan total reward yang diterima seiring pembelajaran. Proses dari Q-learning sendiri adalah sebagai berikut.

   1. Inisialisasi tabel Q untuk menyimpan nilai Q untuk setiap state dan aksi yang dilakukan dengan nilai 0.
   2. Selama pembelajaran, agen akan berperilaku sesuai dengan nilai random yang didapatkan. Jika nilai random tersebut menandakan bahwa agen harus eksplorasi lingkungan pembelajaran, maka aksi dari agen adalah random. Jika nilai random menandakan bahwa agen harus menggunakan informasi yang telah dipelajari, maka agen akan beraksi berdasarkan Q Table.
   3. Setelah agen memilih suatu aksi, Q Table di-update dengan menggunakan rumus berikut.



Dimana nilai dari q value yang baru adalah q value saat ini ditambah dengan learning rate dikali dengan reward dari aksi di state tersebut yang ditambah dengan discount rate dikali dengan reward maksimum ke depannya dan dikurangi dengan q value saat ini.

4. Lakukan langkah di atas hingga seluruh iterasi selesai. Setiap iterasi tentu bisa berbeda karena nilai random yang didapatkan bisa berbeda dan nilai Q Table juga berubah seiring pembelajaran dilakukan.
5. Setelah pembelajaran selesai, Q Table dapat digunakan untuk menentukan keputusan terbaik untuk setiap statenya.

3. Jelaskan perbedaan algoritma Q-learning dan algoritma SARSA, serta kelemahan dan kekurangan masing-masing algoritma!

Perbedaan utama dari kedua algoritma tersebut adalah cara mengupdate nilai dari Q Table.

1. **Q-Learning:**
$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_{t+1} + \gamma max_a Q(s_{t+1}, a) - Q(s_t, a_t))$$
2. **SARSA:**
$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))$$

Q-learning adalah metode off policy dimana ia memperbarui nilai tabel Q berdasarkan tindakan yang optimal secara greedy sedangkan SARSA adalah metode on policy dimana ia memperbarui nilai tabel Q berdasarkan tindakan yang saat ini sedang dilakukan oleh agen.

Kelemahan dari Q-Learning adalah memerlukan waktu yang lama untuk konvergen terutama pada lingkungan dengan state dan aksi yang banyak. Lalu, dengan adanya probabilitas eksplorasi, kemungkinan untuk agen terus bereksplorasi juga berdasarkan probabilitasnya sehingga agen bisa saja sulit atau lama untuk menemukan solusi yang optimal.

Kelemahan dari SARSA adalah memerlukan waktu yang lama untuk konvergen dibandingkan Q-Learning karena pengupdatean Q Table berdasarkan tindakan yang tidak selalu optimal. SARSA juga sensitif terhadap hiperparameter yang ditentukan sehingga bisa saja sulit atau lama juga untuk menemukan solusi yang optimal.