الفرق بين خوارزمية Q-Leaming و Sarsa ا

في مجال التعلم الآلي وتحديدًا التعلم المعزز، تعتبر خوارزميات Q-Learning و Sarsa من أهم الخوارزميات المستخدمة في تدريب الوكلاء الآليين على اتخاذ قرارات استنادًا إلى تفاعلاتهم مع البيئة. كل من هذه الخوارزميات يستخدم أسلوبًا مختلفًا في تحسين سلوك الوكيل بناءً على المكافآت التي يتلقاها من البيئة. على الرغم من تشابه الهدف، إلا أن هناك فروق جوهرية بين الخوارزميتين من حيث الاستراتيجية التي تتبعها كل منهما.

Q-Learning

تُعتبر خوارزمية Q-Learning نوعًا من خوارزميات التعلم غير المرتبط بالسياسة (off-policy)، مما يعني أنها لا تعتمد على السياسة الحالية التي يستخدمها الوكيل لاختيار الأفعال. تعتمد على حساب أفضل سياسة ممكنة بغض النظر عن الأفعال التي ينفذها الوكيل في البيئة. تقوم بتحديث القيم باستخدام معادلة بسيطة تعتمد على الفرق بين القيمة المتوقعة والمكافأة الفعلية التي يتم الحصول عليها.

Sarsa

من ناحية أخرى، تُعتبر خوارزمية Sarsa نوعًا من خوارزميات التعلم المرتبطة بالسياسة (on-policy)، مما يعني أنها تعتمد على السياسة الحالية للوكيل أثناء اتخاذ القرارات. اسم Sarsa يأتي من الحروف الأولى لكل من الحالة (State)، الفعل (Action)، المكافأة (Reward)، الحالة التالية (Next State)، والفعل التالي (Next Action). تتطلب خوارزمية Sarsa اتخاذ القرار التالي حتى تقوم بتحديث القيم.

خاتمة

الفرق الأساسي بين Q-Learning و Sarsa يكمن في كيفية تحديث كل منهما للقيم: Q-Learning لا يعتمد على السياسة الحالية، بينما Sarsa يعتمد عليها. هذا يجعل Sarsa أكثر تحفظًا في قراراتها، بينما تكون Q-Learning أكثر جرأة.