

الفرق بين خوارزمية Q-Learning و Sarsa ٢

تعد خوارزميات التعلم المعزز من الأدوات الفعالة لتعليم الآلات كيفية التصرف بناءً على التجربة. كل من Q-Learning و Sarsa يستخدمان لتحسين أداء الوكلاء بناءً على البيئة، لكن كل خوارزمية تتبع نهجًا مختلفًا في كيفية اتخاذ القرارات وتحديث القيم.

طريقة التحديث في Q-Learning

في خوارزمية Q-Learning، يتم تحديث قيم Q بعد تنفيذ الفعل في البيئة والحصول على مكافأة. يتم حساب التحديث بناءً على الفعل الأفضل الممكن من الحالة التالية. وهذا يعني أن الوكيل يحاول دائماً تحسين سلوكه بأخذ أفضل خيار ممكن في المستقبل، مما يجعل هذه الخوارزمية أكثر ميلاً إلى المخاطرة وأخذ أفعال جريئة.

طريقة التحديث في Sarsa

في المقابل، يتم تحديث القيم في خوارزمية Sarsa بناءً على السياسة الحالية التي يتبعها الوكيل. بدلاً من الاعتماد على أفضل فعل ممكن، يتم تحديث القيم بناءً على الفعل الذي يتخذه الوكيل بالفعل. هذا يجعل Sarsa أكثر تحفظاً في اتخاذ القرارات ويأخذ بعين الاعتبار المخاطر المحتملة بناءً على سياسة الوكيل الحالية.

خاتمة

بينما يُنظر إلى Q-Learning على أنها خوارزمية أكثر طموحاً تأخذ في الاعتبار المستقبل الأكثر تفاعلاً، فإن Sarsa تأخذ نهجاً أكثر واقعية وحذراً يعتمد على الأفعال الحالية. يعتمد اختيار الخوارزمية على طبيعة المهمة والمخاطر التي يرغب المطور في تحملها.

