

شرح ذكاء اصطناعي

التعلم المعزز

اداء/محمد الصلاحي

التعلم المعزز (Reinforcement Learning) هو أحد فروع الذكاء الاصطناعي والتعلم الآلي، حيث يتعلم الوكيل (Agent) كيفية اتخاذ قرارات لتحقيق هدف معين من خلال التفاعل مع البيئة المحيطة به. في التعلم المعزز، يتعلم الوكيل من خلال المحاولة والخطأ، حيث يتم مكافأته على التصرفات الصحيحة ومعاقبته على التصرفات الخاطئة. الهدف الرئيسي هو أن يتعلم الوكيل الاستراتيجية المثلى التي تزيد من مكافأته بمرور الوقت.

1. مكونات التعلم المعزز:

التعلم المعزز يعتمد على ثلاثة مكونات رئيسية:

- الوكيل (Agent): هو الكيان الذي يتخذ القرارات بناءً على التفاعل مع البيئة.
- البيئة (Environment): هي العالم الذي يعمل فيه الوكيل، والذي يحتوي على جميع القواعد والمعلومات حول كيفية استجابة البيئة لتصرفات الوكيل.
- الإجراءات (Actions): هي الخيارات المتاحة للوكيل في كل خطوة، والتي يمكنه اتخاذها للتفاعل مع البيئة.
- الحالة (State): هي تمثيل للبيئة في لحظة معينة، وتحدد الوضع الذي يتواجد فيه الوكيل.
- المكافآت (Rewards): هي التعزيزات التي يحصل عليها الوكيل نتيجة لأفعاله. قد تكون إيجابية أو سلبية، وهي تلعب دوراً مهماً في توجيه التعلم.

2. العملية الأساسية في التعلم المعزز:

التعلم المعزز يعتمد على دورة مستمرة بين الوكيل والبيئة. الخطوات الأساسية هي:

- الحالة الحالية (State): الوكيل يلاحظ حالته الحالية في البيئة.
- اختيار إجراء (Action): الوكيل يختار إجراء بناءً على سياسته الحالية (Policy)، وهي الاستراتيجية التي يتبعها لاتخاذ القرار.
- الحصول على مكافأة (Reward): بعد اتخاذ الإجراء، تقوم البيئة بتقديم مكافأة للوكيل بناءً على ما إذا كان الإجراء يؤدي إلى نتائج جيدة أو سيئة.
- تحديث الحالة (State Transition): البيئة تتغير نتيجة الإجراء الذي اتخذه الوكيل، وينتقل إلى حالة جديدة.
- تحديث السياسة (Policy Update): الوكيل يقوم بتحديث استراتيجيته بناءً على المعلومات الجديدة التي حصل عليها من المكافأة وحالة البيئة.

3. أنواع التعلم المعزز:

هناك نوعان أساسيان من استراتيجيات التعلم المعزز:

أ) التعلم القائم على القيمة: (Value-based Learning)

في هذا النوع، يتعلم الوكيل قيمة كل حالة أو قيمة الإجراء في حالة معينة، ويحاول تعظيم هذه القيمة. أشهر خوارزمية لهذا النوع هي:

- خوارزمية Q-Learning: تعتمد على بناء مصفوفة من الحالات والإجراءات (Q-Table)، حيث يتم تحديث قيم كل إجراء بناءً على المكافأة التي يحصل عليها الوكيل.

ب) التعلم القائم على السياسة: (Policy-based Learning)

في هذا النوع، يتعلم الوكيل السياسة المثلى مباشرةً، وهي عبارة عن دالة تحدد الإجراء الأفضل في كل حالة. يتم استخدام خوارزميات لتحديث هذه السياسة بناءً على المكافآت.

ج) التعلم المختلط: (Actor-Critic Methods)

يجمع هذا النوع بين التعلم القائم على القيمة والتعلم القائم على السياسة، حيث يتعلم الوكيل السياسة المثلى ويقوم أيضاً بتقدير القيم المرتبطة بالحالات والإجراءات.

4. دالة القيمة ودالة Q:

- دالة القيمة (Value Function): هي تقدير للقيمة المتوقعة التي يمكن للوكيل الحصول عليها من حالة معينة، أي مجموع المكافآت المستقبلية التي يمكنه تحقيقها بدءاً من تلك الحالة.
- دالة Q (Q-Function): هي تقدير لقيمة اتخاذ إجراء معين في حالة معينة. تساعد دالة Q في تحديد الإجراء الأمثل في كل حالة.

5. الاستكشاف والاستغلال: (Exploration vs. Exploitation)

من التحديات الرئيسية في التعلم المعزز هو الموازنة بين:

- الاستكشاف (Exploration): اكتشاف إجراءات جديدة لم تجربها الوكيل من قبل، والتي قد تكون أفضل من الإجراءات التي يعتمد عليها حاليًا.
- الاستغلال (Exploitation): استغلال المعرفة الحالية للوكيل لتحديد أفضل إجراء بناءً على المعلومات المتاحة حتى الآن.

عادةً ما تستخدم خوارزميات التعلم المعزز تقنيات مثل سياسة الإيسيلون (ϵ -greedy) لتحقيق التوازن بين الاستكشاف والاستغلال.

6. أمثلة عملية للتعلم المعزز:

أ) الألعاب (Games):

أحد أشهر تطبيقات التعلم المعزز هو في مجال الألعاب. على سبيل المثال، تم استخدام التعلم المعزز لتدريب برنامج AlphaGo الذي طورته شركة DeepMind والذي تمكن من هزيمة أبطال العالم في لعبة "Go". يعتمد AlphaGo على استخدام التعلم المعزز لتعلم كيفية اللعب بشكل مثالي.

ب) الروبوتات (Robotics):

في الروبوتات، يستخدم التعلم المعزز لتدريب الروبوتات على التحكم في حركتها واتخاذ القرارات بشكل مستقل. على سبيل المثال، يتم استخدام التعلم المعزز لتعليم الروبوتات كيفية التنقل في بيئة معقدة وتجنب العقبات.

ج) التداول المالي (Financial Trading):

يستخدم الذكاء الاصطناعي، بما في ذلك التعلم المعزز، في تطوير أنظمة تداول آلية تتعلم كيفية اتخاذ قرارات شراء وبيع الأسهم بناءً على البيانات التاريخية للسوق.

د) السيارات ذاتية القيادة (Autonomous Vehicles):

السيارات ذاتية القيادة تعتمد على التعلم المعزز لتعلم كيفية اتخاذ قرارات القيادة في مواقف مختلفة مثل التعرف على الإشارات المرورية، التعامل مع المشاة، واتخاذ المسارات الآمنة.

7. التحديات في التعلم المعزز:

- مشكلة البيئة المعقدة: قد تكون بعض البيئات معقدة جدًا وتحتوي على عدد كبير من الحالات، مما يجعل من الصعب على الوكيل اكتشاف الاستراتيجيات المثلى.
- الاستكشاف غير الفعال: إذا لم يُمكّن الوكيل باستكشاف البيئة بشكل كافٍ، فقد يفوت فرصًا لتحسين استراتيجيته.
- المكافآت المتأخرة: في بعض الأحيان، قد يحصل الوكيل على مكافأة بعد وقت طويل من اتخاذ الإجراء الصحيح، مما يجعل عملية التعلم أكثر تعقيدًا.

8. في النهاية نلخص الكلام في الآتي:

التعلم المعزز هو تقنية قوية تتيح للوكيل تعلم اتخاذ قرارات مثلى من خلال التفاعل مع البيئة وتلقي التغذية الراجعة. تستخدم هذه التقنية على نطاق واسع في مجالات متعددة مثل الألعاب، الروبوتات، والسيارات ذاتية القيادة. على الرغم من التحديات، فإن التعلم المعزز يمثل خطوة مهمة نحو تطوير أنظمة ذكاء اصطناعي أكثر تطورًا وفعالية في حل المشكلات المعقدة.