

مریم رضانی امیری-۰۵۴۰۸۱۰۵۴۱۶۳۴۱-هوش مصنوعی سکشن ۱۳-استاد عصایی

تمرین ۱: رفتار در مسائل غیر قطعی چگونه است؟

فرایند تصمیم‌گیری مارکوف (MDP) یک چارچوب ریاضیاتی است که برای مدل‌سازی تصمیم‌گیری در محیط‌های تصادفی استفاده می‌شود. MDP ها به طور خاص در یادگیری تقویتی و نظریه کنترل کاربرد دارند. این فرایند شامل مجموعه‌ای از عناصر کلیدی است که در ادامه به آن‌ها اشاره می‌کنم:

| اجزای فرایند تصمیم‌گیری مارکوف:

1. وضعیت‌ها: (States)

- مجموعه‌ای از وضعیت‌ها که سیستم می‌تواند در آن‌ها قرار گیرد. هر وضعیت نمایانگر یک نقطه خاص در محیط است.

2. عملکردها: (Actions)

- مجموعه‌ای از عمل‌هایی که عامل (Agent) می‌تواند در هر وضعیت انجام دهد. انتخاب عمل بستگی به وضعیت فعلی دارد.

3. انتقال وضعیت: (State Transition)

- تابعی که احتمال انتقال از یک وضعیت به وضعیت دیگر را بر اساس عمل انجام شده مشخص می‌کند. این تابع معمولاً به صورت $P(s' | s, a)$ بیان می‌شود، که نشان‌دهنده احتمال انتقال به وضعیت s' از وضعیت s با انجام عمل a است.

4. پاداش: (Reward)

- تابع پاداش که مقدار پاداشی را که عامل در هر انتقال وضعیت دریافت می‌کند، مشخص می‌کند. این تابع معمولاً به صورت $R(s, a)$ تعریف می‌شود.

5. سیاست: (Policy)

- سیاست یک تابع است که به هر وضعیت، یک عمل را نسبت می‌دهد. سیاست می‌تواند قطعی (deterministic) یا تصادفی (stochastic) باشد.

6. هدف: (Objective)

- هدف عامل به حداکثر رساندن مجموع پاداش‌های دریافتی در طول زمان است. این معمولاً با استفاده از یک معیار ارزش‌گذاری مانند ارزش کل پاداش‌های آینده (expected cumulative reward) انجام می‌شود.

| ویژگی‌های فرایند تصمیم‌گیری مارکوف:

- حالت مارکوف: خاصیت اصلی MDP ها این است که آینده فقط به وضعیت فعلی و عمل انتخاب شده بستگی دارد و نه به تاریخچه وضعیت‌ها و اعمال قبلی. این خاصیت به "حالت مارکوف" معروف است.
- پاداش‌های تخفیف‌دار: معمولاً برای محاسبه ارزش پاداش‌های آینده از یک فاکتور تخفیف (γ) استفاده می‌شود تا پاداش‌های دورتر کمتر از پاداش‌های نزدیک‌تر در نظر گرفته شوند.

| حل: MDP

برای حل MDP ها و پیدا کردن سیاست بهینه، می‌توان از روش‌های مختلفی استفاده کرد، از جمله:

- روش‌های ارزشی: مانند الگوریتم ارزش‌گذاری و الگوریتم سیاست‌گذاری.
- روش‌های مبتنی بر سیاست: مانند الگوریتم‌های سیاست‌گذاری مستقیم.
- یادگیری تقویتی: مانند Q-learning و SARSA.

| کاربردها:

- فرایندهای تصمیم‌گیری مارکوف در بسیاری از حوزه‌ها کاربرد دارند، از جمله رباتیک، بازی‌های رایانه‌ای، سیستم‌های کنترل و مدیریت منابع.
- با استفاده از MDP ها، می‌توان مسائل پیچیده تصمیم‌گیری را به صورت ساختاریافته و قابل حل مدل‌سازی کرد.