تمرین چهارم درس هوش مصنوعی

دانشکده مهندسی برق و کامپیوتر

فربد سیاهکلی ۸۱۰۱۹۸۵۱۰

منطق گزاره‌ای

۱) ابتدا گزاره‌های لازمه را به صورت زیر تعریف کرده:

D: سرعت بیشتر از $120\frac{Km}{h}$ است

R: رد شدن از چراغ قرمز

F: جریمه شدن

$$D\ V\ R \rightarrow F$$

۲) ابتدا گزاره‌ها را به صورت زیر تعریف کرده:

W: آخر هفته بودن

R: باران باریدن

C: کوهنوردی رفتن

D: تمام شدن کار علی

H: در خانه فیلم دیدن

$$\big((W \wedge D) \rightarrow C\big) \wedge (R \rightarrow H)$$

۳) ابتدا گزاره‌ها را به صورت زیر تعریف کرده:

S: با پشتکار درس خواندن

H: دریافت نمره بالا

D: تحویل به موقع تکالیف

E: در امتحانات عالی بودن

A: حضور فعال داشتن در کلاس

$$[(S \wedge D) \rightarrow H] \wedge \big[\big(\sim S \vee (E \vee A)\big) \rightarrow \sim H\big]$$

منطق مرتبه اول

۱) هر فردی حداقل یک کار را انجام داده است.

$$\forall x_i \in x\ (\exists y_i \in y \rightarrow\ P(x_i, y_i))$$

فردی وجود دارد که همه کارها را انجام داده است.

$$\exists x_i \in x\ (\forall y_i \in y \rightarrow\ P(x_i, y_i))$$

فردی وجود دارد که هیچ کاری انجام نداده است.

$$\exists x_i \in x\ (\forall y_i \in y \rightarrow \sim P(x_i, y_i))$$

یک کار وجود دارد که توسط همه انجام داده شده باشد

$$\exists y_i \in y\ \ (\forall x_i \in x \rightarrow\ P(x_i, y_i))$$

۲) گزاره‌ها را به صورت زیر تعریف می‌کنیم.

$P(x)$: یعنی x سیاست‌مدار است.

$F(x, y, t)$: یعنی فرد x، y را در زمان t فریب میدهد.

$$(\ [\exists y\ (\forall P(x) \rightarrow F(x, y, t))] \wedge [\forall y\ (\exists t \rightarrow F(x, y, t))]\ ) \Leftrightarrow [\forall y\ (\exists P(x) \rightarrow \sim F(x, y, t))]$$

استنتاج

صحت گزاره Q به سه ترم $P, R, (M \wedge N)$ وابسته است که کافیست یکی از آنها درست باشد.

- **ترم اول** به صورت زیر است:

$$P \rightarrow Q$$

که به روش backward chaining داریم

$$A \wedge B \rightarrow P$$

لذا برای درست بودن Q عبارات A و B نیز هردو باید درست باشند. می‌دانیم عبارت A درست است اما در رابطه با B اطلاعاتی نداریم پس نمی‌توانیم آن را اثبات کنیم.

- به سراغ **ترم دوم** می‌رویم:

$$R \rightarrow Q$$

لذا برای درست بودن Q عبارت R نیز باید درست باشد که در رابطه با R اطلاعاتی نداریم و توانایی اثبات از این مسیر نیز نداریم.

- از **ترم سوم** داریم:

$$M \wedge N \rightarrow Q$$

$$A \rightarrow M, \qquad D \rightarrow N$$

راجب به A و D می‌دانیم که هر دو صحیح هستند. در نتیجه عبارات M و N نیز صحیح هستند و چون برای صحیح بودن Q هردوی M و N باید صحیح باشند، می‌توان اثبات کرد که Q نیز صحیح است.

به این ترتیب با استفاده از روش backward chaining، صحت گزاره‌ی Q بررسی شد.

پروسه تصمیم‌گیری مارکوف

الف)

$$V^*(s) = \max_a Q^*(s,a)$$

$$Q^*(s,a) = \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma V^*(s')\right]$$

$$V^*(s) = \max_a \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma V^*(s')\right]$$

در این مسئله نویز وجود ندارد. ماتریس جایزه $Reward$ به صورت زیر است:

0	0	+100
0	0	0
Initial	0	-100

با ماتریس $Value$ زیر شروع کرده و آنها را هربار آپدیت می‌کنیم. $(\lambda = 0.9)$

0	0	+100
0	0	0
Initial	0	0

می‌دانیم که $agent$ به میزان یکسان دو $action$ بالا و یا راست رفتن را انتخاب می‌کند. ابتدا همسایه‌های ماتریس آبی رنگ آپدیت می‌شود.

0	50	+100
0	0	50
Initial	0	0

در مرحله بعدی برای ماتریس بالا چپ داریم:

X	50	+100
0	Y	50
Initial	0	Z

$$X = \frac{1}{2} \times (0 + \lambda \times X) + \frac{1}{2} \times (0 + \lambda \times 50)$$

$$\rightarrow X = \frac{\lambda X}{2} + \frac{45}{2} \rightarrow X = 40.91$$

که در اینجا ترم اول مربوط به اکشن بالا رفتن است که با انجام اینکار از آنجایی که خانه‌ای در بالای $Agent$ قرار ندارد، در همان خانه می‌ماند و X می‌گیرد.

در ادامه معادله‌ای نسبتا شبیه برای X را برای Z نیز داریم:

$$Z = \frac{1}{2} \times (0 + \lambda \times 50) + \frac{1}{2} \times (0 + \lambda \times Z)$$

$$\rightarrow Z = 40.91$$

برای Y نیز داریم:

$$Y = \frac{1}{2} \times (0 + \lambda \times 50) + \frac{1}{2} \times (0 + \lambda \times 50)$$

$$\rightarrow Y = 45$$

40.91	50	+100
0	45	50
Initial	0	40.91

حال در $iteration$ بعدی دوباره ارزش‌ها را آپدیت می‌کنیم.

40.91	50	+100
X	45	50
Initial	Y	40.91

برای X نیز داریم:

$$X = \frac{1}{2}(0 + \lambda \times 40.91) + \frac{1}{2}(0 + \lambda \times 45) = 38.66$$

برای Y نیز داریم:

$$Y = \frac{1}{2}(0 + \lambda \times 45) + \frac{1}{2}(-100 + \lambda \times 40.91) = -11.35$$

در نتیجه ماتریس نهایی ارزش‌ها به شرح زیر شده است و مسیر نهایی نیز به رنگ سبز درآمده:

40.91	50	+100
38.66	45	50
Initial	-11.35	40.91

ب) ضریب تخفیف در الگوریتم‌های یادگیری تقویتی و برنامه‌ریزی پویا مانند ارزش‌گذاری به‌کار می‌رود تا از تفاوت زمانی پاداش‌ها در نظر گرفته شود. این ضریب نماینده تمایل به پاداش‌های فوری در مقابل پاداش‌های تاخیری است.

چند دلیل برای استفاده از ضریب تخفیف وجود دارد:

عدم قطعیت آینده: ضریب تخفیف به ما امکان می‌دهد شرایطی را که نتایج آینده در آنها ناقطعی هستند، مدل‌سازی کنیم. با کاهش ارزش پاداش‌های آینده، اهمیت کمتری به پاداش‌هایی که در آینده دورتر هستند اختصاص می‌دهیم، زیرا درباره آنها سطح بالاتری از عدم قطعیت وجود دارد.

ترجیح زمانی: به‌طور کلی، انسان‌ها تمایل دارند به پاداش‌های فوری نسبت به پاداش‌های تاخیری. ضریب تخفیف به کمک کاهش ارزش پاداش‌های آینده نسبت به پاداش‌های فوری، این ترجیح زمانی را در نظر می‌گیرد.

همگرایی و پایداری: ضریب تخفیف اطمینان می‌یابد که الگوریتم ارزش‌گذاری به همگرایی می‌رسد و تخمین‌های ارزش پایداری را ارائه می‌دهد. بدون تخفیف، تابع ارزش ممکن است همگرا نشود یا بین ارزش‌های مختلف نوسان کند که سخت‌تر است راهکاری بهینه را پیدا کنیم.

افق محدود در مقابل افق بی‌نهایت: در برخی موارد، ضریب تخفیف برای مدل‌سازی مسائلی با افق زمانی محدود به‌کار می‌رود. با تنظیم ضریب تخفیف به ۰، به‌طور موثر پاداش‌های آینده را فراتر از یک نقطه مشخص نادیده می‌گیریم و فرض می‌کنیم که مسئله یک نقطه پایان محدود دارد.

در کل، ضریب تخفیف کمک می‌کند تا تعادلی بین پاداش‌های فوری و آینده برقرار شود و به ما امکان می‌دهد در یک فرآیند تصمیم‌گیری متوالی با در نظر گرفتن عدم قطعیت و ترجیح زمانی تصمیم‌های بهینه بگیریم.