

حل مسئله‌ی گاوچران با استفاده از روش Q-Learning

صورت مساله

محیط مسئله یک صفحه‌ی ۵*۵ به شکل زیر است. موقعیت عامل‌ها، آغل، مانع‌ها و گاو توسط شکل زیر مشخص شده‌است. هدف مسئله وارد کردن گاو به آغل با استفاده از یادگیری تقویتی چند عاملی است. گاو همواره از حرکت به سمت عامل‌ها فرار می‌کند و عامل‌ها باید گاو را به نحوی هدایت کنند که گاو مجبور به حرکت به سمت آغل شود و وارد آن شود. عامل‌ها از موقعیت آغل و مانع‌ها، اطلاعی ندارند.

				Herder_1(1)
	Obstacle(-1)	Obstacle(-1)		
				Cow(50)
Corral(100)				Herder_2(2)

روش حل

برای حل مسئله دو فایل `server` و `agent` در اختیار شما قرار گرفته است.

:Agent

برای این مسئله نیاز است که یک کلاس برای عامل ایجاد کنید که از کلاس `BaseAgent` ارث بری کند. در این پروژه نیاز به دو عامل دارید. برای ایجاد هر عامل به نام عامل، موقعیت عامل، موقعیت عامل دیگر و مکان گاو نیاز است.

عامل به صورت پیش فرض یک تابع به نام `message_passing` دارد که از طریق آن می تواند از مکان اطلاع دیگر اطلاع پیدا کند.

:Server

سرور، محیط که بالا را شبیه سازی می کند. سرور دارای پنج تابع قابل دسترسی است .

۱. `start_simulation`: در شروع شبیه سازی باید این تابع فراخوانی شود. این تابع ورودی ندارد. خروجی آن به ترتیب مکان عامل اول، مکان عامل دوم، مکان گاو و یک مقدار بولین است که نشان میدهد گاو داخل آغل است یا خیر.

۲. `send_action`: این تابع، اکشن عامل ها را دریافت می کند. ورودی این تابع به ترتیب اکشن انتخاب شده توسط عامل و شماره است و خروجی ندارد.

۳. `step`: پس از اینکه سرور اکشن ها را از عامل ها دریافت کرد، با فراخوانی این تابع؛ سرور یک مرحله شبیه سازی را انجام میدهد و اگر عامل اول یا دوم، اکشن خود را به سرور ارسال نکرده باشند؛ سرور به صورت پیش فرض اکشن عامل را `skip` یا بی حرکت فرض میکند. این تابع ورودی و خروجی ندارد.

۴. `get_precept`: پس از فراخوانی تابع `step`، هر عامل باید تابع `get_precept` را فراخوانی کند تا `state` جدید خود و گاو را دریافت کند. ورودی این تابع، خود عامل است و خروجی ندارد.

۵. `show_states`: با فراخوانی این تابع، یک دید از کل شبکه برای شما چاپ می شود. این تابع ورودی و خروجی ندارد.

هر عامل می تواند به هشت جهت (هشت خانه ی مجاور) حرکت کرده یا در محل خود ساکن بماند. اگر دو عامل بخواهند به یک مکان بروند، سرور به صورت تصادفی یکی را به آن مکان، انتقال می دهد.

گاو در چهار جهت می تواند حرکت کند(اریب نمیتواند حرکت کند) و تا جای ممکن، وارد آغل نمی شود مگر اینکه فضای دیگری وجود نداشته باشد و مجبور شود. گاو به صورت کلی به سمت مکان هایی می رود که اطرافشان خالی باشد و از عامل ها نیز دوری می کند.

قرار است این مساله را با روش Q-Learning حل کنید، پس لازم است که جدول مناسبی برای ذخیره ی state-action ها ایجاد نمایید. پس از انجام فاز یادگیری عامل ها، با انجام فاز تست، میزان یادگیری عامل ها را ارزیابی کنید.

نکات

- مقداردهی reward و پارامتر های دیگر بر عهده ی شماست.
- در انتها نمودار مجموع پاداش ها بر حسب episode ها را ترسیم کنید.
- پیاده سازی گرافیکی نمره اضافی دارد.
- به پیاده سازی های مشابه، نمره ی صفر تعلق می گیرد.