

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۱	مقدمه	۰۱	۱	یادداشت مستقیم

مقدمه ای بر استراتژی تکاملی

استراتژی های تکاملی، خانواده ای از الگوریتم های بهینه سازی است که برای ما مانند جعبه سیاه بنظر میرسد. این الگوریتم برای آموزش شبکه های عصبی عمیق، آموزش Q و روش های گرادیان کاربرد زیادی دارد که الگوریتم های یادگیری تقویتی عمیق را با چالش کشیده و ثابت کرده بسیار سریع تر عمل میکند که به دلیلی موازی بودن آن است.

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۲	مقدمه	۰۱	۲	یادداشت غیر مستقیم

نحوه کارکرد استراتژی تکاملی

استراتژی تکاملی از الگوریتم تکامل طبیعی که در هر مرحله، جمعیتی تشکیل شده و طی فرآیندی پرورش یافته و تکامل پیدا میکند. در آخر این مرحله بخشی از جمعیت جدید بوجود آمده جهش پیدا میکند تا جا برای تغییرات احتمالی که منجر به بهبود یافتن روند شود، وجود داشته باشد. هدف الگوریتم به حداقل/حداکثر رساندن یک هدف والی است که طی این روند بدست می آید. استراتژی تکاملی با تغییراتی، بهتر از این الگوریتم کار کرده و نتایج چشم گیری بدست آورده.

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۳	۱	۰۱	۳	یادداشت غیر مستقیم

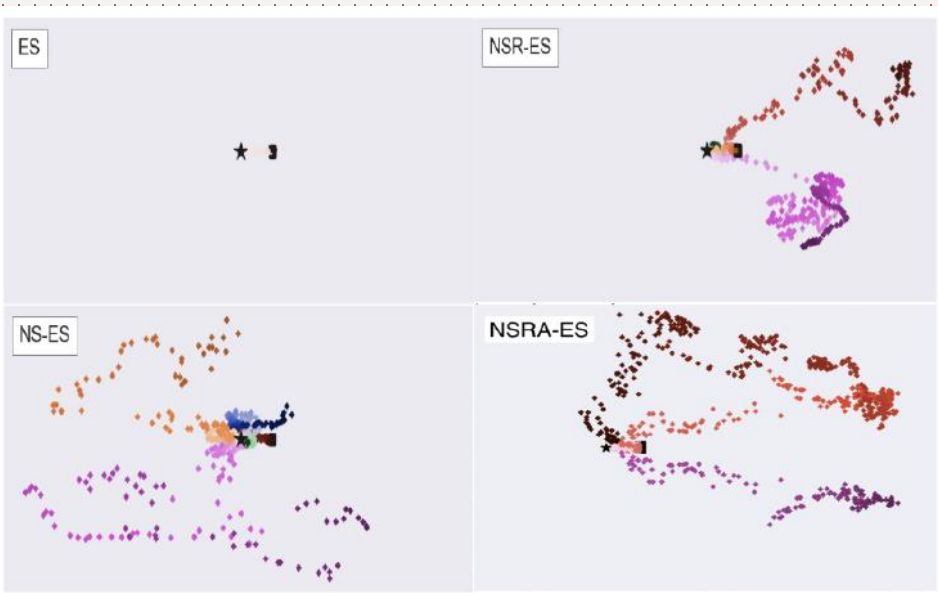
روش ها مورد استفاده قرار گرفته شده برای استراتژی تکاملی

استراتژی تکاملی که حکم یک چهارچوب برای الگوریتم های متعددی در آموزش عامل ها است. از معروف ترین این الگوریتم ها میتوان به NS-ES و QD-ES که خود به دو زیرشاخه دیگه تبدیل میشوند اشاره کرد که دقت و بهینه سازی بالاتری نسبت به دگر الگوریتم ها دارند. هر دو شیوه از شیوه مادر یعنی NOVELTY SEARCH الهام گرفته و پایه ای بر آن اساس دارند.

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۴	۱	۰۱	۵	یادداشت غیر مستقیم

آزمایش: مشکل حرکت انسان نما شبیه سازی شده

این آزمایش به این صورت است که یک انسان نما با ایستادن امتیاز مثبت و با افتادن بر روی زمین امتیاز منفی دریافت میکند. همچنین مقدار انرژی مورد استفاده برای ایستادن از فاکتور های مهم امتیاز دادن است. برای تمام الگوریتم های یادگیری که در این استراتژی صورت گرفت، این آزمایش صورت گرفته تا نتایج مقایسه شده و الگوریتم ها بر اساس بهینگی رتبه بندی شوند. شکل مقابل اهمیت الگوریتم های مطرح شده برای جلوگیری از ماکسیم های محلی را نشان میدهد.



شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۵	۱	۰۱	۷	یادداشت غیر مستقیم

آزمایش: بازی آتاری

هدف این آزمایش مانند آزمایش قبلی (فیش ۴) مقایسه شیوه های یادگیری مورد استفاده در این استراتژی است اما در محیطی جدید و پارامترهای متفاوت زیرا این بازی نیاز به کنترل بیشتری دارد. نتایج حاصل شده از این شبیه سازی نشان میدهد الگوریتم NS مکانیسم چشم گیری برای مواجهه با ماکسیم های محلی ارائه میدهد. این نتیجه نشان میدهد الگوریتم NOVERTY SEEKER میتواند به تنهایی نتیجه بهتری نسبت به دگر الگوریتم ها داده و با ادغام شدن با استراتژی تکاملی دقت آن را به صورت چشمگیری افزایش دهد.

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۶	۱	۰۱	۸	یادداشت غیر مستقیم

مقایسه عملکرد الگوریتم های یادگیری توسط بازی های متعدد

GAME	ES	NS-ES	NSR-ES	NSRA-ES	DQN	NoisyDQN	A3C+
ALIEN	3283.8	1124.5	2186.2	4846.4	2404	2403	1848.3
AMIDAR	322.2	134.7	255.8	305.0	924	1610	964.7
BANK HEIST	140.0	50.0	130.0	152.9	455	1068	991.9
BEAM RIDER [†]	871.7	805.5	876.9	906.4	10564	20793	5992.1
FREEWAY [†]	31.1	22.8	32.3	32.9	31	32	27.3
FROSTBITE [†]	367.4	250.0	2978.6	3785.4	1000	753	506.6
GRAVITAR	1129.4	527.5	732.9	1140.9	366	447	246.02
MONTEZUMA	0.0	0.0	0.0	0.0	2	3	142.5
Ms. PACMAN	4498.0	2252.2	3495.2	5171.0	2674	2722	2380.6
Q*BERT [†]	1075.0	1234.1	1400.0	1350.0	11241	15545	15804.7
SEAQUEST [†]	960.0	1044.5	2329.7	960.0	4163	2282	2274.1
ZAXXON	9885.0	1761.9	6723.3	7303.3	4806	6920	7956.1

برای بررسی دقیق تر و حاصل شدن اطمینان بیشتر نسبت به عملکرد این الگوریتم ها، نیاز است آن ها را در شرایط متفاوت بررسی و نتایج را ثبت کرد. این کار بر روی ۱۲ بازی مختلف انجام شده و نتایج به گونه مقابل میباشد:

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۷	۱	۰۱	۸	یادداشت غیر مستقیم

اکتشاف مبتنی بر جمعیت در مقابل اکتشاف تک عاملی

وجود تعداد متعدد عملگر برای مسائل با حالت های متعدد که نیاز به اکتشاف دارند حیاتی است و باعث میشود سرعت افزایش پیدا کند. در روش های سنتی مانند BFS یا DFS یک عامل باید مسافت زیادی طی کند تا حالات ممکن بررسی شده و بهترین انتخاب شود در صورتی که اگر یک جمعیت برای اینکار بوجود می آمد هر عامل میتواند بررسی های خود را کرده و در اختیار بقیه بگذارد. همچنین با این روش حالت های موفقیت بیشتری تجربه شده و در زمان کمتر عمل بیشتری صورت میگیرد.

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۸	۱	۰۱	۹	یادداشت غیر مستقیم

انتخاب یک ویژگی رفتاری مناسب

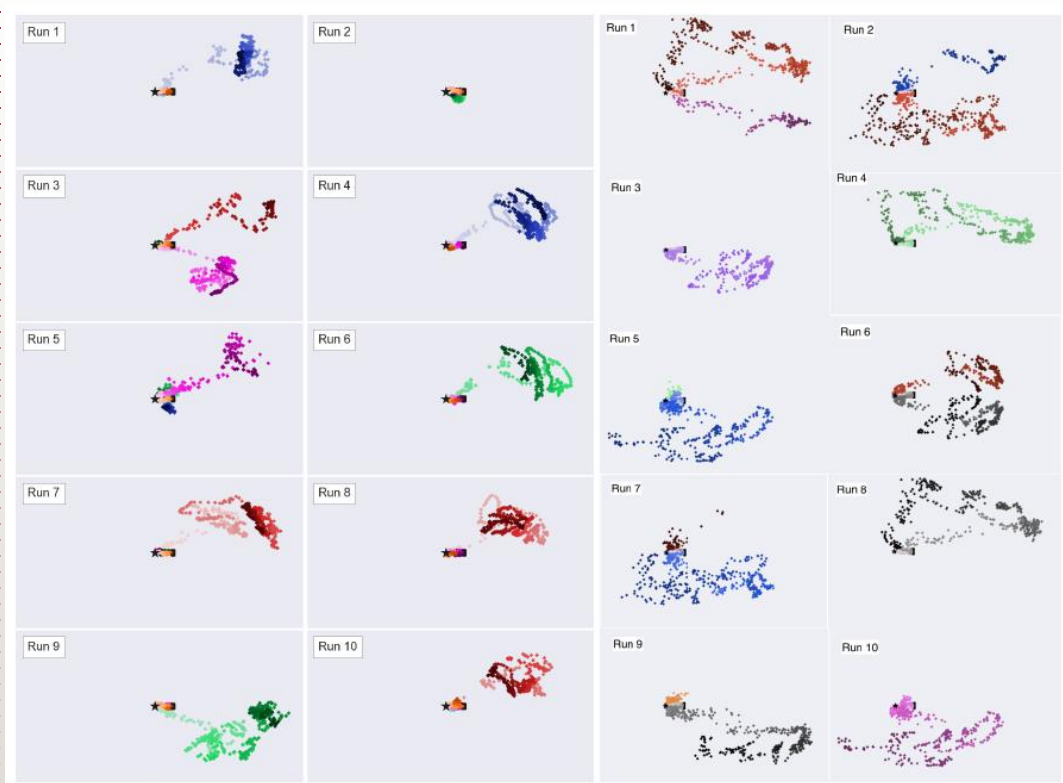
با این وجود که الگوریتم های مورد بحث قرار گرفته بهیچگی بالایی دارند، اما بدون وجود داشتن یک بستر شخصیتی هر عامل میتواند یک فاکتور حیاتی در سریع تر کردن جست و جو و کاهش زیرساخت های مورد نیاز برای انجام عملیات باشد. روش هایی وجود دارد که بستر شخصیت عامل ها را در حین شبیه سازی توسعه داده و تکمیل میکنند (بخش عمیق).

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۹	۱	۰۱	۱۲	یادداشت غیر مستقیم

حفظ مقیاس پذیری

از جنبه های مهم این استراتژی میتوان به توسعه پذیری با سخت افزار موجود اشاره کرد، به طوری که محدودیتی برای مقیاس این چهارچوب وجود ندارد. با وجود اینکه بین الگوریتم های مطرح شده تفاوت هایی وجود داشته و برخی میزان سخت افزار بیشتری نیاز دارند، اما همه قابلیت موازی کار کردن را داشته و نیازی نیست اول هر مرحله جمعیت تکامل پیدا بکند (این روند تدریجی است) باعث کمتر شدن نیاز به سخت افزار قوی میشود.

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
۹	۱	۰۱	۱۵	یادداشت غیر مستقیم



بررسی توان محاسباتی مورد نیاز الگوریتم ها

در راست NSRA-ES و چپ NSR-ES مشاهده میشود. واضح است الگوریتم سمت چپ به توان محاسباتی کمتری نیاز داشته و سریع تر به هدف خود میرسد.

شماره فیش	بخش مربوطه	شماره منبع	صفحه منبع	نوع یادداشت
10	۱	۰۱	۹	یادداشت مستقیم

نتیجه گیری

دو روش NS و QD برای جلوگیری از ماکسیمم های محلی طراحی شده اند و در الگوریتم یادگیری عمیق استفاده بسزایی دارند که نه تنها قابلیت انعطاف پذیری و مقیاس پذیری به این چهارچوب اضافه میکنند، بلکه در تعیین هدف و بهینه سازی آن هم تاثیر بسزایی دارند. همچنین افزایش جمعیت به پیشرفت سریع تر عملیات کمک میکند و در زمان صرفه جویی میشود.