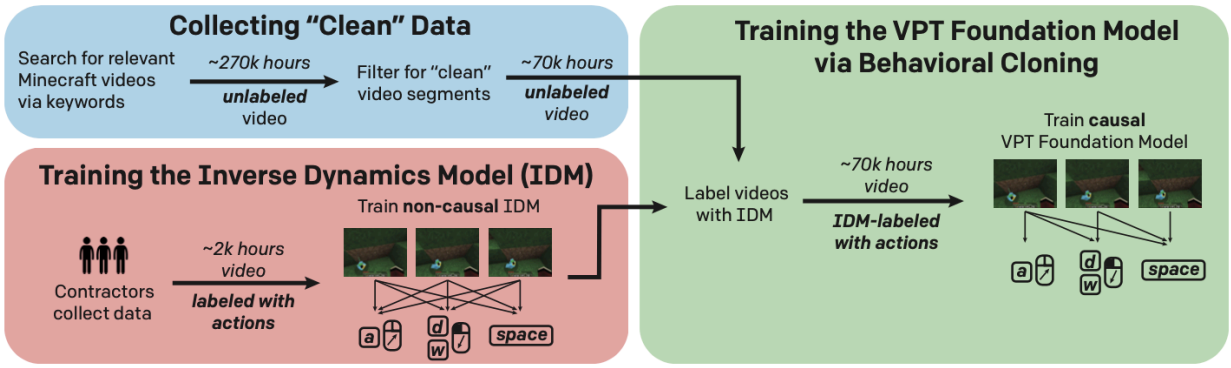
Persian Summary of Article 2206.11795 by Ramtin Kosari

Video PreTraining

انجام عمل به واسطه یادگیری ویدیو های طبقه بندی نشده



چکیده

پیش یادگیری روی مجموعه داده های اینترنتی با محتواهای گوناگون و پراکنده، روشی شناخته شده برای آموزش مدل های هوش مصنوعی در حوزه های مختلف مانند متن و تصویر است. با این حال، در محیط های تصمیم گیری متوالی مانند رباتیک و بازی‌های ویدیویی، داده های طبقه بندی شده شده ممکن است همیشه به آسانی در دسترس نباشند. برای مقابله با این موضوع، این مقاله رویکرد جدیدی به نام یادگیری تقلیدی نیمه نظارتی را معرفی می‌کند که در آن عامل های هوش مصنوعی با مشاهده ویدیو های آنلاین بدون طبقه بندی یاد می‌گیرند. با استفاده از مقدار کمی از داده‌های طبقه بندی شده، می توانیم مدلی را برای طبقه بندی دقیق منبع وسیعی از داده های طبقه بندی نشده، مانند ویدیوهای افرادی که Minecraft بازی می‌کنند، آموزش دهیم. این مدل آموزش دیده میتواند کارهای پیچیده را با حداقل نظارت انجام دهد و حتی می تواند در سناریوهای خاص از انسان ها بهتر عمل کند. علاوه بر این، ما نشان می دهیم که این مدل را می توان با یادگیری تقویتی برای مقابله با وظایف چالش برانگیز تنظیم کرد و تطبیق پذیری و کارایی آن را به نمایش گذاشت. این کار نشان دهنده پیشرفت قابل توجهی در زمینه هوش مصنوعی است.

پروژه پیش یادگیری شرکت OpenAI

شرکت OpenAI یک شبکه عصبی برای بازی Minecraft توسط پیش یادگیری در یک مجموعه داده عظیم ویدیویی بدون طبقه بندی از بازی Minecraft که انسان ها به عنوان بازی کننده قرار دارند با فقط مقدار کمی از داده های طبقه بندی شده ساخته است. این مدل شبکه عصبی می تواند یاد بگیرد که ابزار الماس در بازی درست کند، کاری که معمولاً انسان های ماهر معمولا در حدود ۲۰ دقیقه انجام میدهند. این مدل شبکه عصبی از محیط بازی مشابه با انسان استفاده میکند و قابلیت های وارد کردن کلید کیبورد و همچنین حرکت موس را نیز دارد. خود این مورد بیانگر قدمی به سوی عامل های استفاده کننده از کامپیوتر است

چرا Minecraft

شرکت OpenAI تصمیم گرفته متود خود را در Minecraft پیاده سازی کند زیرا (۱) یکی از پرطرفدار ترین بازی های ویدیویی در جهان است و بنابراین انبوهی از داده های ویدیویی مجانی را در دسترس قرار میدهد و (۲) یک بازی پایان باز است و مانند دنیای واقعی میتوان هر کاری را انجام داد، مانند استفاده از کامپیوتر در دنیای واقعی و بازی کردن

مدل بنیادین

این مدل بر روی 70000 ساعت ویدیوی آنلاین با برچسب IDM آموزش داده شده، مدل شبیه سازی رفتاری ("مدل بنیادین VPT") وظایفی را در Minecraft انجام می دهد که دستیابی به آنها با یادگیری تقویتی از ابتدا تقریبا غیر ممکن است. می آموزد که درختان را خرد کند تا کنده ها را جمع کند، آن کنده ها را به صورت تخته درآورد، و سپس آن تخته ها را در یک میز کار استفاده کند. این سکانس تقریباً 50 ثانیه یا 1000 اکشن متوالی بازی برای یک انسان حرفه ای در Minecraft طول می کشد.

بهبود دقت با شبیه سازی رفتاری

مدل‌های بنیادین به گونه‌ای طراحی شده‌اند که نمایه رفتاری گسترده‌ای داشته باشند و به طور کلی در طیف گسترده‌ای از وظایف قادر باشند. برای ترکیب دانش جدید یا اجازه دادن به آنها برای تخصص در توزیع وظایف محدودتر، دقت این مدل ها را با مجموعه داده های کوچکتر و خاص تر بهبود میدهند. به عنوان یک مطالعه خاص در مورد اینکه چگونه می توان مدل پایه VPT را به خوبی برای مجموعه داده های پایین دستی تنظیم کرد، از بازی کنندگان خود خواستیم که به مدت 10 دقیقه در دنیای جدید Minecraft بازی کنند و خانه ای از مواد اولیه Minecraft بسازند. ما امیدوار بودیم که این توانایی مدل بنیادین برای اجرای بهینه مهارت های "شروع بازی" مانند ساخت میز کار را تقویت کند.

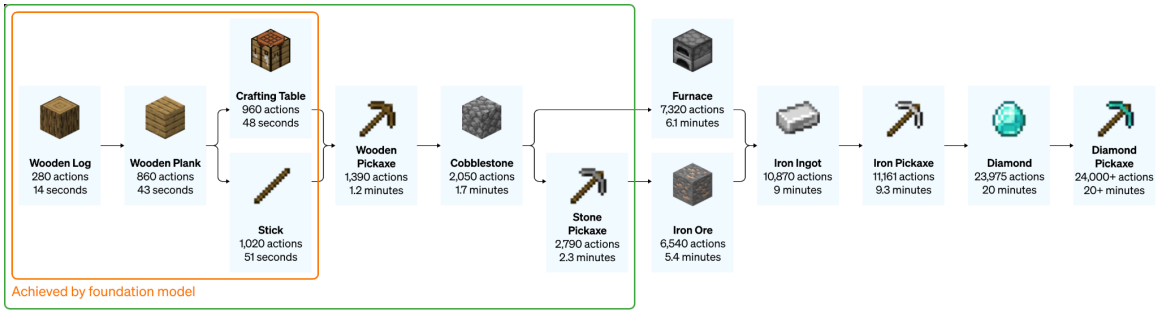
بهبود دقت با یادگیری تقویتی

هنگامی که امکان تعریف یک تابع پاداش وجود دارد، یادگیری تقویتی (RL) می تواند یک روش قدرتمند برای استخراج عملکرد بالا، حتی بالقوه فوق انسانی باشد. در عین حال بسیاری از وظایف نیاز بر اکتشافات سخت دارند و اکثر روش‌های RL با اولویت‌های اکتشاف و با غلبه بر تصمیمات تصادفی کار میکنند. مدل VPT باید برای RL فواید بسیار بهتری داشته باشد زیرا شبیه سازی رفتار انسان در ابتدا احتمالاً بسیار مفیدتر از انجام اقدامات تصادفی است.

شرکت OpenAI وظیفه چالش برانگیز مدل خود را جمع آوری کلنگ الماسی قرار داده، قابلیتی بی سابقه در Minecraft که هنگام استفاده از رابط کاربری انسانی دشوارتر می شود.

ساخت کلنگ الماسی مستلزم یک سری وظایف فرعی طولانی و پیچیده است. برای اینکه این کار قابل انجام باشد، برای هر مورد به ترتیب به عوامل پاداش می دهیم.

ما دریافتیم که RL که از یک مقداردهی اولیه تصادفی آموزش داده شده است (روش استاندارد RL) و ب ندرت به پاداشی دست پیدا میکند، هرگز یاد نمی گیرد که تنه درختان را جمع آوری کند و فقط به ندرت چوب ها را جمع آوری می کند. در مقابل، بهبود دقت یک مدل VPT نه تنها یاد می‌گیرد که کلنگ‌های الماسی بسازد (که در 2.5 درصد ویدیو های 10 دقیقه‌ای Minecraft انجام می‌شود)، بلکه حتی در جمع‌آوری همه موارد منجر به ساخت کلنگ الماسی نیز موفق میشود. این اولین باری است که کسی به یک عامل کامپیوتری نشان داده که قادر به ساخت ابزارهای الماس در Minecraft است که به طور متوسط بیش از 20 دقیقه (24000 عمل) برای انسان زمان می برد.



نتیجه گیری

VPT مسیری را هموار می کند که به عوامل اجازه می دهد تا با تماشای تعداد زیادی ویدیو در اینترنت عمل ما را انجام دهد. در مقایسه با مدل‌سازی ویدیویی تولیدی یا روش‌های متضاد که فقط مقدمات بازنمایی را ارائه می‌دهند، VPT امکان هیجان‌انگیز یادگیری مستقیم مقدمات رفتاری در مقیاس بزرگ را در حوزه های بیشتری به جای زبان ارائه می‌دهد. در حالی که ما فقط در Minecraft آزمایش می کنیم، بازی بسیار باز است و رابط انسانی (موس و صفحه کلید) استفاده میشود، بنابراین شرکت OpenAI معتقد است که نتایج شان برای سایر دامنه های مشابه خوب است.