

مدل‌های زبانی بزرگ به کمک ربات‌ها می‌آیند

یک روش جدید هدایت ربات‌ها، از ورودی‌های مبتنی بر زبان به جای داده‌های بصری پرهزینه استفاده می‌کند و این کار را در چند مرحله انجام می‌دهد.



یک روش جدید هدایت ربات‌ها، از ورودی‌های مبتنی بر زبان به جای داده‌های بصری پرهزینه استفاده می‌کند و این کار را در چند مرحله انجام می‌دهد.

به گزارش ایسنا، ممکن است روزی بخواهید ربات خانگی شما لباس‌های کثیف را به طبقه پایین خانه ببرد و آنها را در ماشین لباسشویی بگذارد. ربات باید دستورالعمل‌های شما را با مشاهدات بصری خود ترکیب کند تا مراحل تکمیل این کار را تشخیص دهد.

به نقل از ساینس دیلی، شاید یک روش جدید بتواند مسیری را برای ربات تنها با استفاده از ورودی‌های مبتنی بر زبان برنامه‌ریزی کند. اگرچه این روش نمی‌تواند بهتر از روش‌های مبتنی بر بینش رایانه‌ای عمل کند اما می‌تواند در تنظیماتی که فاقد داده‌های بصری برای آموزش دادن ربات هستند، سودمند باشد.

برای یک عامل هوش مصنوعی، حرف زدن درباره این کار آسان‌تر از انجام دادن آن است. روش‌های کنونی اغلب از چندین مدل یادگیری ماشینی برای برخورد با بخش‌های گوناگون کار استفاده می‌کنند که به تلاش و تخصص انسانی نیاز دارند. این روش‌ها که از بازنمایی‌های بصری به منظور تصمیم‌گیری مستقیم برای جهت‌یابی استفاده می‌کنند، حجم زیادی از داده‌های بصری را برای آموزش می‌خواهند که اغلب به سختی به دست می‌آیند.

پژوهشگران دانشگاه «ام آی تی» (MIT) و «آزمایشگاه هوش مصنوعی واتسون ام آی تی-آی بی ام» (MIT-IBM Watson AI Lab) برای غلبه بر این چالش‌ها، یک روش جهت‌یابی ابداع کردند که نمایش‌های بصری را به بخش‌های زبانی تبدیل می‌کند. سپس این بخش‌ها به یک مدل زبانی بزرگ وارد می‌شوند که همه بخش‌های وظیفه جهت‌یابی چندمرحله‌ای را انجام می‌دهد.

روش آنها به جای رمزگذاری ویژگی‌های بصری، از تصاویر محیط اطراف ربات به عنوان نمایش‌های بصری استفاده می‌کنند که از نظر محاسباتی فشرده است و زیرنویس‌های متنی ارائه می‌دهد که نقطه نظر ربات را توصیف می‌کنند. یک مدل زبانی بزرگ از این توصیف‌ها، برای پیش‌بینی اقداماتی استفاده می‌کند که یک ربات باید برای انجام دادن دستورالعمل‌های مبتنی بر زبان کاربر انجام دهد.

از آنجا که روش این گروه پژوهشی از بازنمایی‌های صرفاً مبتنی بر زبان استفاده می‌کند، آنها می‌توانند یک مدل زبانی بزرگ را برای تولید موثر حجم زیادی از داده‌های آموزش مصنوعی به کار بگیرند.

اگرچه این روش نسبت به روش‌هایی که از ویژگی‌های بصری استفاده می‌کنند، عملکرد بهتری ندارد اما در موقعیت‌هایی که فاقد داده‌های بصری کافی برای آموزش هستند، به خوبی عمل می‌کند. پژوهشگران دریافتند که ترکیب ورودی‌های مبتنی بر زبان با سیگنال‌های بصری، به عملکرد بهتر در جهت‌یابی کمک می‌کند.

«بوون پن» (Bowen Pan) پژوهشگر ارشد این پروژه گفت: روش ما با استفاده صرف از زبان به عنوان نمایش ادراکی، ساده‌تر عمل می‌کند. از آنجا که همه ورودی‌ها را می‌توان به صورت زبان رمزگذاری کرد، می‌توانیم یک مسیر قابل درک را برای انسان ایجاد کنیم.