



دانشگاه تهران

دانشکده علوم و فنون نوین

گزارش هفته اول

فاطمه چیت ساز	نام و نام خانوادگی
830402092	شماره دانشجویی
25 آذر 1402	تاریخ ارسال گزارش

- 1.....Foundation Models مطالعه بر روی
- 3.....مطالعه عمیق تر بر روی مقالات مربوطه

مطالعه بر روی Foundation Models

مقدمه:

در این گزارش، به بررسی و تحلیل مدل‌های اساسی (Foundation Models) اخیراً معرفی شده می‌پردازیم. این مدل‌ها از توانایی‌ها و قابلیت‌های بالایی برخوردارند و در حوزه‌های مختلفی از جمله تولید متن، دید کامپیوتری، تصویرسازی متن به تصویر، و سایر زمینه‌ها به کار گرفته می‌شوند.

***Florence**

مدل Florence با استفاده از جفت‌های متن به تصویر آموزش دیده است و در حال حاضر به عنوان یک سرویس تجاری آماده برای دید کامپیوتری در Azure Cognitive Service for Vision یکپارچه شده است. این مدل قابلیت‌هایی از جمله تجزیه و تحلیل تصاویر، خواندن متن و تشخیص چهره با برچسب‌گذاری تصویر از پیش ساخته را دارد و امکان سفارشی‌سازی با داده خود را نیز فراهم می‌کند. Florence یک مدل چندحالتی (multimodal) است که به تحلیل متن، تصویر، و حتی ویدیو می‌پردازد.

لینک مرتبط : (<https://arxiv.org/pdf/2303.10130.pdf>)

***Llama**

Llama یک مدل منبع‌باز است که با تأکید بر حریم خصوصی فراهم شده است. این مدل امکان تنظیم دقیق (fine-tuning) با زبان انگلیسی را دارد و از ۷ میلیارد تا ۷۰ میلیارد پارامتر (7B, 13B, 70B) پشتیبانی می‌کند. نسخه‌های پیش‌آموزش‌دیده این مدل دارای بهبودهای قابل ملاحظه‌ای می‌باشند و از ۴۰٪ توکن بیشتری آموزش دیده شده‌اند، طول متن بیشتری (۴k توکن) دارند.

لینک مرتبط: (<https://arxiv.org/pdf/2307.09288.pdf>)

***GPT (Generative Pre-trained Transformer)**

GPT-4 با یک تریلیون پارامتر، جلوه‌ای از افتخارات اخیر این خانواده مدل‌هاست. این مدل ابتدا یک پیش‌آموزش بی‌نظارت را با استفاده از یک مجموعه بزرگ از داده‌ها طی می‌کند، سپس با تنظیم دقیق تحت نظر آموزش میبند تا به مدل در انجام وظایف خاص کمک کند. GPT-3 همچنین با بیش از ۱۷۵ میلیارد پارامتر یکی از نسل‌های پیشین این خانواده است.

لازم به ذکر است که GPT-3 حداکثر ۲۰۴۸ توکن را به عنوان محدودیت ورودی دارد.

DALL-E

DALL-E یک مدل پیشرفته است که از تکنیک‌هایی چون Discrete Variational Auto-Encoder (dVAE) و Denoising Diffusion Probabilistic Models برای تولید تصاویر متناظر با متن ورودی استفاده می‌کند. این مدل همچنین با استفاده از مدل clip ارزیابی انطباق بین تصاویر و متن را انجام می‌دهد.

Bard

Bard هدف از امکان پرسش‌ها و دستورات کاربر به صورت زبان طبیعی است، به جای استفاده از کلمات کلیدی برای جستجو. پاسخ‌های ChatGPT براساس داده‌های تا آوریل ۲۰۲۳ ایجاد شده‌اند، درحالی که Google Bard بر اساس اطلاعات به روز خواهد بود.

ارتباط با BERT: BERT وظیفه بهبود نتایج جستجوی گوگل را دارد و به صورت پنهانی عمل می‌کند. اما Bard، آخرین پروژه هوش مصنوعی گوگل است و به سوالات و دستورات کاربر پاسخ می‌دهد.

ابزارها:

دو ابزار معروف در زمینه مدل‌های اساسی عبارتند از Lang Chain و Hugging Face. Hugging Face یک شرکت است که یک پلتفرم برای به اشتراک گذاری و همکاری در مدل‌های یادگیری ماشین فراهم می‌کند.

یکی از ویژگی‌های محبوب Hugging Face، model hub است که مجموعه بزرگی از مدل‌های پیش‌آموزش‌دیده در اختیار قرار می‌دهد.

Hugging Face همچنین ابزارهای متعددی برای آموزش و ارزیابی مدل‌های پردازش زبان طبیعی ارائه می‌دهد. این ابزارها شامل موارد زیر می‌شوند:

- Transformers: یک کتابخانه برای آموزش و ارزیابی مدل‌های مبتنی بر ترنسفورمر
- Datasets: یک مجموعه از مجموعه‌داده‌ها برای پژوهش‌های NLP
- Metrics: یک مجموعه از معیارها برای ارزیابی مدل‌های پردازش زبان طبیعی
- LangChain: یک چارچوب طراحی شده است تا ایجاد برنامه‌ها با استفاده از مدل‌های زبان بزرگ (LLMs) را ساده‌تر کند.

منابع مفید:

برای مطالعه بیشتر می‌توانید به لینک‌های زیر مراجعه کنید:

- (<https://arxiv.org/pdf/2108.07258.pdf>)

- (<https://www.techtarget.com/whatis/feature/Foundation-models-explained-Everything-you-need-to-know>)

مطالعه عمیق تر بر روی مقالات مربوطه

مقدمه

در ادامه برای درک مقالات مجبور به مطالعه یک سری مسائل مقدماتی در ماشین لرنینگ شدم که در ادامه به بررسی تعدادی از آنان می‌پردازیم

Recurrent Neural Networks

شبکه‌های عصبی بازگشتی (RNNs) نوعی از شبکه‌های عصبی هستند که می‌توانند از داده‌های دنباله‌ای یاد بگیرند. این ویژگی آنها را برای وظایفی مانند پیش‌بینی بازار، جایگاه مناسبی می‌دهد، زیرا قیمت یک سهم در یک زمان مشخص به قیمت‌های زمان‌های گذشته وابسته است.

یکی از چالش‌ها در استفاده از شبکه‌های عصبی بازگشتی برای پیش‌بینی بازار این است که گرادیان‌ها ممکن است به صورت افزایشی یا ناپدید شونده در طول زمان عمل کنند. این ممکن است باعث سختی آموزش شبکه و عملکرد ناپایدار شود.

vanishing

گرادیان‌های ناپدید شونده زمانی رخ می‌دهند که گرادیان‌های تابع خطا نسبت به وزن‌های شبکه به مقدار بسیار کوچک می‌شوند. این ممکن است زمانی رخ دهد که وزن‌ها به مقادیر کوچکی مقداردهی اولیه شوند یا زمانی که توابع فعال‌سازی در شبکه شیبی دیگر از صفر نداشته باشند.

زمانی که گرادیان‌ها ناپدید می‌شوند، شبکه قادر به یادگیری از داده نمی‌شود و فرآیند آموزش متوقف می‌شود.

exploding gradient

گرادیان‌های افزایشی زمانی رخ می‌دهند که گرادیان‌های تابع خطا نسبت به وزن‌های شبکه به مقدار بسیار بزرگ می‌شوند. این ممکن است زمانی رخ دهد که وزن‌ها به مقادیر بزرگی مقداردهی اولیه شوند یا زمانی که توابع فعال‌سازی در شبکه شیبی به یک نزدیک باشند.

زمانی که گرادیان‌ها افزایش می‌یابند، شبکه ناپایدار می‌شود و فرآیند آموزش ممکن است انحراف پیدا کند. راهکار: (LSTM):

LSTM به عنوان یک راهکار برای چالش‌های مرتبط با گرادیان‌های ناپدید و افزایشی معرفی می‌شوند. ساختار دوگانه: واحدهای LSTM از دو مسیر تشکیل شده‌اند؛ مسیر حافظه کوتاه‌مدت که ورودی فعلی را با یک وزن در نظر می‌گیرد و مسیر حافظه طولانی مدت که از وزن استفاده نمی‌کند.

توابع فعال‌سازی: از توابع فعال‌سازی sigmoid و tanh استفاده می‌شود که تبدیلات غیرخطی فراهم کرده و به شبکه اجازه می‌دهد که ارتباطات پیچیده را در داده‌های دنباله‌ای ثبت کند.

مدیریت حافظه: عملیات ضرب (*) برای تعیین درصد حافظه طولانی مدتی که باید یادآوری شود، استفاده می‌شود